

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:15:59 ; Search time 20.1923 Seconds
(without alignments)
63.918 Million cell updates/sec

Title: US-10-008-524A-4

Perfect score: 142
Sequence: 1 RPIKPSFWAPKRRRLSDQDSQTP 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	142	100.0	25	4	US-09-314-268-4
2	121	85.2	92	4	US-09-314-268-90
3	110.5	77.8	30	2	US-08-934-915-42
4	98.5	69.4	20	2	US-08-934-915-40
5	98	69.0	35	4	US-09-314-268-117
6	93	65.5	20	2	US-08-934-915-170
7	91	64.1	15	4	US-09-314-268-167
8	75	52.8	12	4	US-09-314-268-168
9	59.5	41.9	39	4	US-09-314-268-119
10	57	40.1	34	4	US-09-314-268-118
11	56.5	39.8	38	4	US-09-314-268-120
12	56.5	39.8	49	4	US-09-314-268-93
13	56	39.4	31	4	US-09-314-268-111
14	56	39.4	32	4	US-09-314-268-114
15	55.5	39.1	442	4	US-09-252-991A-31743
16	54	38.0	173	4	US-09-252-991A-32711
17	53	37.3	143	4	US-09-252-991A-21367
18	53	37.3	202	4	US-09-252-991A-20104
19	52.5	37.0	182	4	US-09-252-991A-27985
20	52	36.6	8	4	US-09-314-268-31
21	52	36.6	8	4	US-09-314-268-33
22	52	36.6	8	4	US-09-314-268-37
23	52	36.6	184	4	US-09-552-322-11
24	51.5	36.3	117	4	US-09-314-268-92
25	51	35.9	8	4	US-09-314-268-36
26	51	35.9	201	3	US-08-987-418A-2
27	51	35.9	201	3	US-09-343-062-2

28 51 35.9 227 6 5498499-2 Patent No. 5498499
29 51 35.9 342 4 US-09-489-039A-9668 Sequence 9668, Ap
30 51 35.9 505 1 US-07-745-206A-23 Sequence 23, Appl
31 51 35.9 505 2 US-08-311-363-23 Sequence 23, Appl
32 51 35.9 993 1 US-08-444-792-2 Sequence 2, Appl
33 51 35.9 993 1 US-08-445-042-2 Sequence 2, Appl
34 51 35.9 1039 4 US-09-409-648-7 Sequence 7, Appl
35 51 35.9 1039 6 5196511-2 Patent No. 5196511
36 50 35.2 8 4 US-09-314-268-34 Sequence 34, Appl
37 50 35.2 8 4 US-09-314-268-35 Sequence 35, Appl
38 50 35.2 8 4 US-09-314-268-38 Sequence 38, Appl
39 50 35.2 144 4 US-09-489-039A-12273 Sequence 12273, A
40 50 35.2 156 4 US-09-252-991A-18484 Sequence 18484, A
41 50 35.2 188 4 US-09-252-991A-19129 Sequence 19129, A
42 50 35.2 195 4 US-09-252-991A-22162 Sequence 22162, A
43 50 35.2 656 4 US-09-252-991A-25003 Sequence 25003, A
44 49.5 34.9 39 4 US-09-314-268-105 Sequence 105, App
45 49 34.5 8 4 US-09-314-268-32 Sequence 32, Appl

ALIGNMENTS

RESULT 1

US-09-314-268-4

; Sequence 4, Application US/09314268

; Patent No. 6346377

; GENERAL INFORMATION:

; APPLICANT: Doorbar, John

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

; TITLE OF INVENTION: VIRUSES

; FILE REFERENCE: 3789/80902

; CURRENT APPLICATION NUMBER: US/09/314.268

; CURRENT FILING DATE: 1999-03-19

; EARLIER APPLICATION NUMBER: 09/314.268

; EARLIER FILING DATE: 1999-05-18

; NUMBER OF SEQ ID NOS: 179

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 25

; TYPE: PRT

; ORGANISM: Human papillomavirus type 16

US-09-314-268-4

Query Match 100.0%; Score 142; DB 4; Length 25;
Best Local Similarity 100.0%; Pred. No. 3.3e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIKPSFWAPKRRRLSDQDSQTP 25

Db 1 RPIKPSFWAPKRRRLSDQDSQTP 25

RESULT 2

US-09-314-268-90

; Sequence 90, Application US/09314268

; Patent No. 6346377

; GENERAL INFORMATION:

; APPLICANT: Doorbar, John

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

; TITLE OF INVENTION: VIRUSES

; FILE REFERENCE: 3789/80902

; CURRENT APPLICATION NUMBER: US/09/314.268

; CURRENT FILING DATE: 1999-03-19

; EARLIER APPLICATION NUMBER: 09/314.268

; EARLIER FILING DATE: 1999-05-18

; NUMBER OF SEQ ID NOS: 179

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 90

; LENGTH: 92

; TYPE: PRT

; ORGANISM: Human papillomavirus type 16

US-09-314-268-90

Query Match 85.2%; Score 121; DB 4; Length 92;
 Best Local Similarity 92.6%; Pred. No. 1.1e-09;
 Matches 25; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
 QY 1 RPIKPSWAPKKHRL-SDQD-SQTP 25
 Db 26 RPIKPSWAPKKHRLSSDQDSQTP 52

RESULT 3
 US-08-934-915-42
 ; Sequence 42, Application US/08934915
 ; Patent No. 5932412
 ; GENERAL INFORMATION:
 ; APPLICANT: DILLNER, JOAKIM
 ; APPLICANT: DILLNER, LENA
 ; APPLICANT: CHENG, HWE-MING
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
 ; TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,
 ; TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
 ; TITLE OF INVENTION: USEFUL IN IMMUNOASSAY FOR
 ; TITLE OF INVENTION: DIAGNOSTIC PURPOSES
 ; NUMBER OF SEQUENCES: 193
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MASON & ASSOCIATES, P.A.
 ; STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
 ; CITY: CLEARWATER
 ; STATE: FLORIDA
 ; COUNTRY: U.S.A.
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: Windows 3.0
 ; SOFTWARE: Microsoft Word 6.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/934,915
 ; FILING DATE: 22-SEP-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/949,836
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: LOUISE A. Foutch
 ; REGISTRATION NUMBER: 37,133
 ; REFERENCE/DOCKET NUMBER: 1946.6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 813-538-3800
 ; TELEFAX: 813-538-3820
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 42:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 30 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-934-915-42

Query Match 77.8%; Score 110.5; DB 2; Length 30;
 Best Local Similarity 95.5%; Pred. No. 8.8e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 RPIKPSWAPKKHRL-SDQD 21
 Db 7 RPIKPSWAPKKHRLSSDQD 28

RESULT 4
 US-08-934-915-40
 ; Sequence 40, Application US/08934915
 ; Patent No. 5932412
 ; GENERAL INFORMATION:
 ; APPLICANT: DILLNER, JOAKIM

; APPLICANT: DILLNER, LENA
 ; APPLICANT: CHENG, HWE-MING
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
 ; TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,
 ; TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
 ; TITLE OF INVENTION: USEFUL IN IMMUNOASSAY FOR
 ; TITLE OF INVENTION: DIAGNOSTIC PURPOSES
 ; NUMBER OF SEQUENCES: 193
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MASON & ASSOCIATES, P.A.
 ; STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
 ; CITY: CLEARWATER
 ; STATE: FLORIDA
 ; COUNTRY: U.S.A.
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: Windows 3.0
 ; SOFTWARE: Microsoft Word 6.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/934,915
 ; FILING DATE: 22-SEP-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/949,836
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: LOUISE A. Foutch
 ; REGISTRATION NUMBER: 37,133
 ; REFERENCE/DOCKET NUMBER: 1946.6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 813-538-3800
 ; TELEFAX: 813-538-3820
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 40:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 20 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-934-915-40

Query Match 69.4%; Score 98.5; DB 2; Length 20;
 Best Local Similarity 95.0%; Pred. No. 2.5e-07;
 Matches 19; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 3 IPKPSWAPKKHRL-SDQD 21
 Db 1 IPKPSWAPKKHRLSSDQD 20

RESULT 5
 US-09-314-268-117
 ; Sequence 117, Application US/09314268
 ; Patent No. 6346377
 ; GENERAL INFORMATION:
 ; APPLICANT: Doorbar, John
 ; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
 ; TITLE OF INVENTION: VIRUSES
 ; FILE REFERENCE: 3789/80902
 ; CURRENT APPLICATION NUMBER: US/09/314,268
 ; CURRENT FILING DATE: 1999-03-19
 ; EARLIER APPLICATION NUMBER: 09/314,268
 ; EARLIER FILING DATE: 1999-05-18
 ; NUMBER OF SEQ ID NOS: 173
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 117
 ; LENGTH: 35
 ; TYPE: PRT
 ; ORGANISM: Human papillomavirus type 16
 ; US-09-314-268-117

Query Match 69.0%; Score 98; DB 4; Length 35;

```
Best Local Similarity 91.3%; Pred. No. 5.5e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 5 KPSPWAPKKHRL-SDQD-SQTP 25
   |||||
Db 1 KPSPWAPKKHRLSSDQSQTP 23
   |||||

RESULT 6
US-08-934-915-170
; Sequence 170, Application US/08934915
; Patent No. 5932412
; GENERAL INFORMATION:
; APPLICANT: DILLNER, JOAKIM
; APPLICANT: DILLNER, LENA
; APPLICANT: CHENG, HWEE-MING
; TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
; TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,
; TITLE OF INVENTION: 11, 15, 18, 31, 33 AND 56,
; TITLE OF INVENTION: USEFUL IN IMMUNOASSAY FOR
; TITLE OF INVENTION: DIAGNOSTIC PURPOSES
; NUMBER OF SEQUENCES: 193
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MASON & ASSOCIATES, P.A.
; STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
; CITY: CLEARWATER
; STATE: FLORIDA
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 3.0
; SOFTWARE: Microsoft Word 6.0
; APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,915
; FILING DATE: 22-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/949,836
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: LOUISE A. Foutch
; REGISTRATION NUMBER: 37,133
; REFERENCE/DOCKET NUMBER: 1946.6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 813-538-3800
; TELEFAX: 813-538-3820
; TELEX:
; INFORMATION FOR SEQ ID NO: 170:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-934-915-170

Query Match 65.5%; Score 93; DB 2; Length 20;
Best Local Similarity 80.0%; Pred. No. 1.5e-06;
Matches 16; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 IPKPSWAPKKHRLSDQDS 22
   |||||
Db 1 IPKPSWAPKKHRLSSQN 20
   |||||

RESULT 7
US-09-314-268-167
; Sequence 167, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES

Best Local Similarity 91.3%; Pred. No. 5.5e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 5 KPSPWAPKKHRL-SDQD-SQTP 25
   |||||
Db 1 KPSPWAPKKHRLSSDQSQTP 23
   |||||

RESULT 8
US-09-314-268-168
; Sequence 168, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 168
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-09-314-268-168

Query Match 52.8%; Score 75; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00025;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PKPSPWAPKKHR 15
   |||||
Db 1 PKPSPWAPKKHR 12
   |||||

RESULT 9
US-09-314-268-119
; Sequence 119, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 119
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Human papillomavirus type 31
US-09-314-268-119
```

Query Match 41.9%; Score 59.5; DB 4; Length 39;
 Best Local Similarity 46.7%; Pred. No. 0.12;
 Matches 14; Conservative 4; Mismatches 3; Indels 9; Gaps 2;

QY 5 KPSPWAPK-----HRRLSDDQ-----SQTP 25
 ||:||||| | |||||: |||
 Db 1 KPAPWAPVKVCGRRRLSDQSQSQSTETP 30

RESULT 10

US-09-314-268-118
 ; Sequence 118, Application US/09314268
 ; Patent No. 6346377
 ; GENERAL INFORMATION:
 ; APPLICANT: Doorbar, John
 ; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
 ; FILE REFERENCE: 3789/80902
 ; CURRENT FILING DATE: 1999-03-19
 ; EARLIER APPLICATION NUMBER: US/09/314,268
 ; EARLIER FILING DATE: 1999-05-18
 ; NUMBER OF SEQ ID NOS: 179
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 118
 ; LENGTH: 34
 ; TYPE: PRT
 ; ORGANISM: Human papillomavirus type 35
 US-09-314-268-118

Query Match 40.1%; Score 57; DB 4; Length 34;
 Best Local Similarity 58.8%; Pred. No. 0.24;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 KPSPWAPKHRRLSDDQ 21
 ||:||||| | |||||
 Db 1 KPAPWAPQPRRQITND 17

RESULT 11

US-09-314-268-120
 ; Sequence 120, Application US/09314268
 ; Patent No. 6346377
 ; GENERAL INFORMATION:
 ; APPLICANT: Doorbar, John
 ; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
 ; FILE REFERENCE: 3789/80902
 ; CURRENT FILING DATE: 1999-03-19
 ; EARLIER APPLICATION NUMBER: US/09/314,268
 ; EARLIER FILING DATE: 1999-05-18
 ; NUMBER OF SEQ ID NOS: 179
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 120
 ; LENGTH: 38
 ; TYPE: PRT
 ; ORGANISM: Human papillomavirus type 52
 US-09-314-268-120

Query Match 39.8%; Score 56.5; DB 4; Length 38;
 Best Local Similarity 52.2%; Pred. No. 0.31;
 Matches 12; Conservative 0; Mismatches 6; Indels 5; Gaps 1;

QY 8 PWAPK-----KHRRLSDDQSQTP 25
 ||||| | |||||
 Db 4 PWVPEKTYNHRNDDQTSQTP 26

RESULT 12

US-09-314-268-93
 ; Sequence 93, Application US/09314268
 ; Patent No. 6346377

; GENERAL INFORMATION:
 ; APPLICANT: Doorbar, John
 ; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
 ; FILE REFERENCE: 3789/80902
 ; CURRENT FILING DATE: 1999-03-19
 ; EARLIER APPLICATION NUMBER: US/09/314,268
 ; EARLIER FILING DATE: 1999-05-18
 ; NUMBER OF SEQ ID NOS: 179
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 93
 ; LENGTH: 49
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: HPV E4 protein
 ; OTHER INFORMATION: consensus amino acid sequence, amino acids 27-73.
 US-09-314-268-93

Query Match 39.8%; Score 56.5; DB 4; Length 49;
 Best Local Similarity 43.8%; Pred. No. 0.41;
 Matches 14; Conservative 2; Mismatches 7; Indels 9; Gaps 2;

QY 2 PIPKSPWA-----PKHRRLL-SDQDSQT 24
 ||||| | |||||
 Db 1 PPPPRPWAGPKPTRGPPRRRLSDSDSDS 32

RESULT 13

US-09-314-268-111
 ; Sequence 111, Application US/09314268
 ; Patent No. 6346377
 ; GENERAL INFORMATION:
 ; APPLICANT: Doorbar, John
 ; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
 ; FILE REFERENCE: 3789/80902
 ; CURRENT FILING DATE: 1999-03-19
 ; EARLIER APPLICATION NUMBER: US/09/314,268
 ; EARLIER FILING DATE: 1999-05-18
 ; NUMBER OF SEQ ID NOS: 179
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 111
 ; LENGTH: 31
 ; TYPE: PRT
 ; ORGANISM: Human papillomavirus type 45
 US-09-314-268-111

Query Match 39.4%; Score 56; DB 4; Length 31;
 Best Local Similarity 54.5%; Pred. No. 0.29;
 Matches 12; Conservative 2; Mismatches 4; Indels 4; Gaps 1;

QY 5 KPSPWAPK-----KHRRLSDDQ 22
 ||||| | |||||
 Db 1 KPSPWAPQNTSRRLSLDLS 22

RESULT 14

US-09-314-268-114
 ; Sequence 114, Application US/09314268
 ; Patent No. 6346377
 ; GENERAL INFORMATION:
 ; APPLICANT: Doorbar, John
 ; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
 ; FILE REFERENCE: 3789/80902
 ; CURRENT FILING DATE: 1999-03-19
 ; EARLIER APPLICATION NUMBER: US/09/314,268
 ; EARLIER FILING DATE: 1999-05-18
 ; NUMBER OF SEQ ID NOS: 179

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 114
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Human papillomavirus type 59
US-09-314-268-114

Query Match      39.4%; Score 56; DB 4; Length 32;
Best Local Similarity 63.6%; Pred. No. 0.3;
Matches 14; Conservative 1; Mismatches 3; Indels 4; Gaps 2;

Qy      5 KPSPWAPKK---HRRLL-SDQDS 22
      |||:|||||
Db      1 KPRTWAPKRGTVRRRLSDQDS 22

RESULT 15
US-09-252-991A-31743
; Sequence 31743, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31743
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31743

Query Match      39.1%; Score 55.5; DB 4; Length 442;
Best Local Similarity 48.0%; Pred. No. 6.2;
Matches 12; Conservative 1; Mismatches 11; Indels 1; Gaps 1;

Qy      1 RPIPKPSPWAPKKHRRLLSDQDSQTP 25
      |||:|||||
Db      138 RPFHGRGRQPHRRR-GDQGRQVP 161

Search completed: May 27, 2004, 16:22:36
Job time : 20.1923 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:21:04 ; Search time 52.8846 Seconds
(without alignments)
132.009 Million cell updates/sec

Title: US-10-008-524A-4

Perfect score: 142

Sequence: 1 RPIPKSPWAPKXRRRLSDQDSQTP 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1151071 seqs, 279249464 residues

Total number of hits satisfying chosen parameters: 1151071

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142	100.0	25	14	US-10-008-524A-4
2	131.5	92.6	26	15	US-10-350-719-4
3	121	85.2	92	14	US-10-008-524A-90
4	121	85.2	92	15	US-10-350-719-90
5	98	69.0	35	14	US-10-008-524A-117
6	98	69.0	35	15	US-10-350-719-117
7	91	64.1	15	14	US-10-008-524A-167
8	82	57.7	15	15	US-10-350-719-167
9	75	52.8	12	14	US-10-008-524A-168
10	75	52.8	12	15	US-10-350-719-168
11	59.5	41.9	39	14	US-10-008-524A-119
12	59.5	41.9	39	15	US-10-350-719-119
13	57	40.1	34	14	US-10-008-524A-118
14	57	40.1	34	15	US-10-350-719-118
15	57	40.1	230	14	US-10-080-170-152

16	56.5	39.8	38	14	US-10-008-524A-120	Sequence 120, App
17	56.5	39.8	38	15	US-10-350-719-120	Sequence 120, App
18	56.5	39.8	49	14	US-10-008-524A-93	Sequence 93, Appl
19	56.5	39.8	49	15	US-10-350-719-93	Sequence 93, Appl
20	56.5	39.8	113	12	US-10-424-599-254084	Sequence 254084,
21	56	39.4	31	14	US-10-008-524A-111	Sequence 111, App
22	56	39.4	31	15	US-10-350-719-111	Sequence 111, App
23	56	39.4	32	14	US-10-008-524A-114	Sequence 114, App
24	56	39.4	32	15	US-10-350-719-114	Sequence 114, App
25	55.5	39.1	495	10	US-09-834-434-3	Sequence 3, Appli
26	55	38.7	119	12	US-10-424-599-250127	Sequence 250127,
27	54	38.0	285	12	US-10-425-114-46504	Sequence 46504, A
28	53	37.3	55	14	US-10-029-386-31681	Sequence 31681, A
29	52.5	37.0	394	11	US-09-833-245-1498	Sequence 1498, Ap
30	52.5	37.0	1590	14	US-10-180-326-1	Sequence 1, Appli
31	52	36.6	8	14	US-10-008-524A-31	Sequence 31, Appl
32	52	36.6	8	14	US-10-008-524A-33	Sequence 33, Appl
33	52	36.6	8	14	US-10-008-524A-37	Sequence 37, Appl
34	52	36.6	8	15	US-10-350-719-31	Sequence 31, Appl
35	52	36.6	8	15	US-10-350-719-33	Sequence 33, Appl
36	52	36.6	8	15	US-10-350-719-37	Sequence 37, Appl
37	52	36.6	184	14	US-10-137-473-11	Sequence 11, Appl
38	52	36.6	184	15	US-10-294-006-14	Sequence 14, Appl
39	52	36.6	196	12	US-10-276-774-1939	Sequence 1939, Ap
40	52	36.6	196	15	US-10-294-006-36	Sequence 36, Appl
41	52	36.6	318	14	US-10-156-761-9265	Sequence 9265, Ap
42	51.5	36.3	117	14	US-10-008-524A-92	Sequence 92, Appl
43	51.5	36.3	117	15	US-10-350-719-92	Sequence 92, Appl
44	51	35.9	8	14	US-10-008-524A-36	Sequence 36, Appl
45	51	35.9	8	15	US-10-350-719-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1

US-10-008-524A-4
; Sequence 4, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-4

Query Match 100.0%; Score 142; DB 14; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIPKSPWAPKXRRRLSDQDSQTP 25

Db 1 RPIPKSPWAPKXRRRLSDQDSQTP 25

RESULT 2

US-10-350-719-4
; Sequence 4, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES

```

; FILE REFERENCE: 13396/2152
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
; US-10-350-719-4

```

Query Match 92.6%; Score 131.5; DB 15; Length 26;
Best Local Similarity 96.2%; Pred. No. 2.2e-09;
Matches 25; Conservative 0; Mismatches 0; Indels 1

Qy	1	RPIKPSPWAPKKHRL-SDQDSQTP	25
Db	1	RPIKPSPWAPKKHRLSSDQDSQTP	26

```

RESULT 3
US/-008-524A-90
, Sequence 90, Application US/10008524A
, Publication No. US20030175692A1
, GENERAL INFORMATION:
, APPLICANT: Doobar, John
, TITLE OF INVENTION: IMPROVEMENTS IN OR REL
, TITLE OF INVENTION: VIRUSES
, FILE REFERENCE: 18396/1074
, CURRENT APPLICATION NUMBER: US/10/008,524A
, CURRENT FILING DATE: 2002-08-13
, PRIOR APPLICATION NUMBER: 09/314,268
, PRIOR FILING DATE: 1999-05-18
, NUMBER OF SEQ ID NOS: 179
, SOFTWARE: Patent in Ver. 2.1
, SEQ ID NO 90
, LENGTH: 92
, TYPE: PRT
, ORGANISM: Human papillomavirus type 16
US/-008-524A-90

```

```
Query Match      85.2%; Score 121; DB 14; Length 92;
Best Local Similarity 92.6%; Pred. No. 1.4e-07;
Matches 25; Conservative 0; Mismatches 0; Indels
```

QY 1 RPIKPSPWAPKKHRL-SDQD-SQTP 25
|||
pb 26 RPIKPSPWAPKKHRL-SSDQDOSQTP 52

```

RESULT 4
US-10-350-719-90
; Sequence 90, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/21162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; LENGTH: 92

```

```

; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-350-719-90

```

Query Match 85.2%; Score 121; DB 15; Length 92;
Best Local Similarity 92.6%; Pred. No. 1.4E-07;
Matches 25; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 RPIKPSPWAPKHHRL-SDQD-SQTP 25
| | | | | | | | | |
db 26 RPIKPSPWAPKHHRLSSDDOSQTP 52

```

RESULT 5
US-10-008-524A-117
? Sequence 117, Application US/10008524A
? Publication No. US20030175692A1
? GENERAL INFORMATION:
? APPLICANT: Doorbar, John
? TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
? TITLE OF INVENTION: VIRUSES
? FILE REFERENCE: 18396/1074
? CURRENT APPLICATION NUMBER: US/10/008,524A
? CURRENT FILING DATE: 2002-08-13
? PRIOR APPLICATION NUMBER: 09/314,268
? PRIOR FILING DATE: 1999-05-18
? NUMBER OF SEQ ID NOS: 179
? SOFTWARE: PatentIn ver. 2.1
? SEQ ID NO 117
? LENGTH: 35
? TYPE: PRT
? ORGANISM: Human papillomavirus type 16
US-10-008-524A-117

```

Query Match 69.0%; Score 98; DB 14; Length 35;
Best Local Similarity 91.3%; Pred. No. 3.9e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY	5	KPSPWAPKXRRLL-SDQD-SQTP	25
Db	1	KPSPWAPKXRRLLSSDQDQSQTP	23

```

RESULT 6
US-10-350-719-117
; Sequence 117, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 117
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-350-719-117

```

Query Match	69.0%	Score 98;	DB 15;	Length 35;
Best Local Similarity	91.3%	Pred. No. 3.9E-05;		
Matches 21; Conservative	0;	Mismatches 0;	Indels 2;	Gaps 2;

QY

5 KPSPWAPKKHRL-SDQD-SQTP 25
|||||

D6

1 KPSPWAPKKHRRLLSSDDQSQT 23
|||||


```
RESULT 7
US-10-008-524A-167
; Sequence 167, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 167
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-167

Query Match      64.1%; Score 91; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIPKPSFWAPKKHR 15
| | | | | | | | | | | | | | |
Db 1 RPIPKPSFWAPKKHR 15

RESULT 8
US-10-350-719-167
; Sequence 167, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 167
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-350-719-167

Query Match      57.7%; Score 82; DB 15; Length 15;
Best Local Similarity 93.3%; Pred. No. 0.0017;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPIPKPSFWAPKKHR 15
| | | | | | | | | | | | | | |
Db 1 RRIKPSFWAPKKHR 15

RESULT 9
US-10-008-524A-168
; Sequence 168, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
```

```
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 168
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-168

Query Match      52.8%; Score 75; DB 14; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0099;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PKPSPWAPKKHR 15
| | | | | | | | | | | | | | |
Db 1 PKPSPWAPKKHR 12

RESULT 10
US-10-350-719-168
; Sequence 168, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 168
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-350-719-168

Query Match      52.8%; Score 75; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0099;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PKPSPWAPKKHR 15
| | | | | | | | | | | | | | |
Db 1 PKPSPWAPKKHR 12

RESULT 11
US-10-008-524A-119
; Sequence 119, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 119
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Human papillomavirus type 31
US-10-008-524A-119
```

Query Match	41.9%;	Score 59.5;	DB 14;	Length 39;
Best Local Similarity	46.7%;	Pred. No. 2.4;		
Matches 14;	Conservative	4;	Mismatches 3;	Indels 9;
				Gaps 2;

Qy 5 KPSPWPKK-----HRLSDQD---SQTP 25
||:||||| | ||||: :||
Db 1 KPAPWAPVKVCGRRRLSDQEQSQSTETP 30

RESULT 12

```

US-10-350-719-119
; Sequence 119, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 119
; LENGTH: 39
; TYPE: FRT
; ORGANISM: Human papillomavirus type 31
US-10-350-719-119

```

Qy 5 KPSWAPKK-----HRLSDQD---SQTP 25
||:||||| | ||||: |||
Db 1 KPAPWAPVKVCGRRRLSDQEQSQSTETP 30

RESULT 13

```

US-10-008-524A-118
; Sequence 118, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 118
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Human papillomavirus type 35
US-10-008-524A-118

```

Qy 5 KPSPWAPKKHRRSLSDQD 21
|||:|||||:|||||
Db 1 KPAPWAPQKPRROITND 17

RESULT 14

US-10-350-719-118
; Sequence 118, Application US/10350719
; Publication No. US2003021972A1
; GENERAL INFORMATION:

; CURRENT APPLICATION NUMBER: US/10/350,719

;
PRIOR I
- UNITED

```

; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 118
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Human papillomavirus type 35
US-10-350-719-118

Query Match          40.1%; Score 57; DB 15; Length 34;
Best Local Similarity 58.8%; Pred. No. 4.3;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 5 KPSPWAPKKHRRLSDDQ 21
   ||:||||:|
Db 1 KPAPWAPQKPRQITND 17

RESULT 15
US-10-080-170-152
; Sequence 152, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 152
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-152

```

Qy 4 FKPSWAPKKHRRRLSDQDS 22
 |||||: ||: ||
 Db 110 FKPSWAPRPLARLLNGDS 128

Search completed: May 27, 2004, 16:34:18
Job time : 53.8846 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:13:48 ; Search time 16.8269 Seconds
(without alignments)
142.913 Million cell updates/sec

Title: US-10-008-524A-4
Perfect score: 142
Sequence: 1 RPIPKSPWAPKRRRLSDQSQTP 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	85.2	95	1 W4WLHS	E4 protein - human
2	80	56.3	96	1 W4WL35	E4 protein - human
3	74	52.1	87	1 W4WL51	E4 protein - human
4	70.5	49.6	102	1 W4WL31	E4 protein - human
5	61.5	43.3	1553	2 T03301	rab3 effector prot
6	59	41.5	94	1 W4WL39	E4 protein - human
7	57	40.1	230	2 S72714	Lepbil70_F2_64 pro
8	56	39.4	137	2 T43028	HMG-protein I alph
9	54	38.0	330	2 T05717	probable extensin
10	53	37.3	88	1 W4WL18	E4 protein - human
11	53	37.3	238	2 T40820	proline-rich prote
12	53	37.3	275	2 A36264	DNA-directed RNA p
13	53	37.3	910	2 UC4609	chitin synthase (E
14	52.5	37.0	684	2 H96646	hypothetical prote
15	52.5	37.0	684	2 T02149	hypothetical prote
16	52	36.6	459	2 S03116	gene 33 protein, h
17	52	36.6	1223	2 T17345	hypothetical prote
18	51	35.9	101	2 S30970	gene 25 protein -
19	51	35.9	1039	2 A34269	integrin alpha-2b
20	51	35.9	1060	2 A10201	beta-galactosidase
21	50.5	35.6	327	2 T45143	hypothetical prote
22	50	35.2	38	1 HSPY5	histone H5 - pigeo
23	50	35.2	105	1 NSCHH4	nonhistone chromos
24	50	35.2	338	2 C75459	probable endonucle
25	49.5	34.9	529	2 A47082	probable transcrip
26	49.5	34.9	529	2 A12119	transcription regu
27	49.5	34.9	838	2 T04785	hypothetical prote
28	49.5	34.9	1487	2 T02850	hypothetical prote
29	49	34.5	101	2 G72802	SP25 protein - Myc

ALIGNMENTS

RESULT 1

W4WLHS

E4 protein - human papillomavirus type 16

C:Species: human papillomavirus type 16

C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 20-Aug-1999

C:Accession: A22355; T10425

R:Seedorf, K.; Krammer, G.; Durst, M.; Suhai, S.; Rowekamp, W.G.

Virology 145, 181-185, 1985

A:Title: Human papillomavirus type 16 DNA sequence.

A:Reference number: A22355; MUID:85246220; PMID:2990099

A:Accession: A22355

A:Molecule type: DNA

A:Residues: 1-95 <SER>

A:Cross-references: GB:K02718; NID:G333031; PIDN:AAA46937.1; PID:G459913

R:Kennedy, I.M.; Haddock, J.K.; Clements, J.B.

J. Virol. 65, 2093-2097, 1991

A:Title: A negative element in the human poapillomavirus type 16 genome acts at the lev

A:Reference number: Z17014; MUID:91162763; PMID:1848319

A:Accession: T10425

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-95 <KEN>

A:Cross-references: EMBL:K02718; NID:G333031; PIDN:AAA46937.1; PID:G459913

C:Genetics:

A:Gene: E4

C:Superfamily: papillomavirus E4 protein

C:Keywords: early protein

Query Match 85.2%; Score 121; DB 1; Length 95;

Best Local Similarity 92.6%; Pred. No. 4.6e-09;

Matches 25; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 RPIPKSPWAPKRRRL-SDQ-SQTP 25

Db 29 RPIPKSPWAPKRRRLSDQSQTP 55

RESULT 2

W4WL35

E4 protein - human papillomavirus type 35

C:Species: human papillomavirus type 35

A:Note: host Homo sapiens (man)

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 27-Jan-1995

C:Accession: C40824

R:Marich, J.E.; Pontsler, A.V.; Rice, S.M.; McGraw, K.A.; Dubensky, T.W.

Virology 186, 770-776, 1992

A:Title: The phylogenetic relationship and complete nucleotide sequence of human papill.

A:Reference number: A40824; MUID:92124753; PMID:1310198

A:Accession: C40824

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-96 <MAR>

A;Cross-references: GB:M74117
C;Superfamily: papillomavirus E4 protein
C;Keywords: early protein

Query Match 56.3%; Score 80; DB 1; Length 96;
Best Local Similarity 66.7%; Pred. No. 0.0012;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RPIKPSWPAPKKHRLSDQD 21
|||||:|||||:|||||
Db 29 RPIKPAWAPQKERRQTND 49

RESULT 3

W4WL51

E4 protein - human papillomavirus type 51

C;Species: human papillomavirus type 51

A;Note: host Homo sapiens (man)

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 27-Jan-1995

C;Accession: C40415

R;Lungu, O.; Crum, C.P.; Silverstein, S.J.

J. Virol. 65, 4216-4225, 1991

A;Title: Biologic properties and nucleotide sequence analysis of human papillomavirus type 51

A;Reference number: A40415; MUID:91303675; PMID:1649326

A;Accession: C40415

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-87 <LUN>

A;Cross-references: GB:M62877

C;Superfamily: papillomavirus E4 protein

C;Keywords: early protein

Query Match 52.1%; Score 74; DB 1; Length 87;
Best Local Similarity 63.6%; Pred. No. 0.0065;
Matches 14; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 RPIKPSWPAPKKHRLSDQD 22
|||||:|||||:|||||
Db 24 RPIPLPAPWAPKPRHNSD 45

RESULT 4

W4WL31

E4 protein - human papillomavirus type 31

C;Species: human papillomavirus type 31

A;Note: host Homo sapiens (man)

C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999

C;Accession: E32444

R;Goldsbrough, M.D.; DiSilvestre, D.; Temple, G.F.; Lorincz, A.T.

Virol. 171, 306-311, 1989

A;Title: Nucleotide sequence of human papillomavirus type 31: a cervical neoplasia-associated virus

A;Reference number: A94398; MUID:89299478; PMID:2545036

A;Accession: E32444

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-102 <GOL>

A;Cross-references: GB:J04353; NID:G333048; PIDN:AAA46949.1; PID:G459915

A;Note: in GenBank entry PP31A the initiation codon UUG for residue 1 is translated as

C;Superfamily: papillomavirus E4 protein

C;Keywords: early protein

Query Match 49.6%; Score 70.5; DB 1; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.022;
Matches 16; Conservative 4; Mismatches 3; Indels 9; Gaps 2;

Qy 3 IPKPSWPAPKK-----HRLSDQD-----SQTP 25
|||||:|||||:|||||
Db 32 IPKPAWAPVVCGRRLSDQSQSTETP 63

RESULT 5

T03301

rab3 effector protein Rim - rat

C;Species: Rattus norvegicus (Norway rat)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C;Accession: T03301
R;Wang, Y.; Okamoto, M.; Schmitz, F.; Hofmann, K.; Sudhof, T.C.

Nature 388, 593-598, 1997

A;Title: Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion.

A;Reference number: Z14897; MUID:97394473; PMID:9252191

A;Accession: T03301

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1553 <WAN>

A;Cross-references: EMBL:AF007836; NID:G2317777; PIDN:AAB66703.1; PID:G2317778

A;Experimental source: tissue-type brain

C;Genetics:

A;Note: RIM

C;Function:

A;Description: Rim protein is proposed as Rab3-dependent regulator of synaptic-vesicle

C;Keywords: GTP binding; zinc finger

Query Match 43.3%; Score 61.5; DB 2; Length 1553;
Best Local Similarity 33.3%; Pred. No. 5.1;
Matches 11; Conservative 7; Mismatches 2; Indels 13; Gaps 1;

Qy 2 PIPKPSWPAPKK-----RLSDQD 21
|||||:|||||:|||||
Db 890 PLQPSFMPRRHNGESSKKLQSRISDSD 922

RESULT 6

W4WL39

E4 protein - human papillomavirus type 39

C;Species: human papillomavirus type 39

A;Note: host Homo sapiens (man)

C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 27-Jan-1995

C;Accession: E38502

R;Volpers, C.; Strebeck, R.E.

Virol. 181, 419-423, 1991

A;Title: Genome organization and nucleotide sequence of human papillomavirus type 39.

A;Reference number: A38502; MUID:91135017; PMID:1847266

A;Accession: E38502

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-94 <VOL>

A;Cross-references: EMBL:M38185

C;Superfamily: papillomavirus E4 protein

C;Keywords: early protein

Query Match 41.5%; Score 59; DB 1; Length 94;
Best Local Similarity 58.1%; Pred. No. 0.66;
Matches 18; Conservative 1; Mismatches 6; Indels 6; Gaps 3;

Qy 1 RPIKPSWPAPKK--HRL-SDQD---SQTP 25
|||||:|||||:|||||
Db 29 RPIPPQHPAPKQSRRLSDLSVQSQP 59

RESULT 7

S72714

Lepb1170_F2_64 protein - Mycobacterium leprae

C;Species: Mycobacterium leprae

C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C;Accession: S72714

R;Smith, D.R.; Robison, K.

submitted to the EMBL Data Library, November 1993

A;Description: Mycobacterium leprae cosmid B1170.

A;Reference number: S72693

A;Accession: S72714

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-230 <SMI>

A;Cross-references: EMBL:U00010; NID:G466780; PIDN:AAA17078.1; PID:G466802

C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rvi517

```

Query Match          40.1%; Score 57; DB 2; Length 230;
Best Local Similarity 57.9%; Pred. No. 2.9;
Matches 11; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy      4 PKPSPWAPKHKRRRLSDQDS 22
      ||||| |||: |||: |||
Db     110 PKPSKWAPRPLARLLNGDS 128

RESULT 8
T43028
A:Title: HMG-protein I alpha chain - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T43028; T16801
R:Kurcz, T.; Schulze, E.
A:Description: The high mobility group proteins of Caenorhabditis elegans.
A:Reference number: Z2282
A:Accession: T43028
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-137 <KUR>
A:Cross-references: EMBL:AF056578; PIDN:AAC78600.1
R:Chisoe, S.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T05A7.
A:Reference number: Z18580
A:Accession: T16801
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-137 <CHI>
A:Cross-references: EMBL:U40028; NID:gl055143; PID:gl055146; PIDN:AAA81116.1; CESP:T05A7
C:Genetics:
A:Gene: hmg-I-alpha; T05A7.4
A:Introns: 29/1

Query Match          39.4%; Score 56; DB 2; Length 137;
Best Local Similarity 44.0%; Pred. No. 2.4;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy      1 RPIKPSWAPKHKRRRLSDQDSQTP 25
      :||| |||: |||: |||: |||
Db     53 KPAPKPYSGAPRGPRKSDQVAKSP 77

RESULT 9
T05717
A:Title: Probable extensin - barley (fragment)
C:Species: Hordeum vulgare (barley)
C:Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 20-Jun-2000
C:Accession: T05717
R:Doan, D.N.P.; Sturaro, M.; Olsen, O.A.
submitted to the EMBL Data Library, July 1997
A:Description: Characterization of a nuclear cDNA encoding a probable extensin from dev
A:Reference number: Z15429
A:Accession: T05717
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-330 <DOA>
A:Cross-references: EMBL:Z98204; PIDN:CAB10894.1
C:Genetics:
A:Gene: ex1
A:Map position: 2
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match          38.0%; Score 54; DB 2; Length 330;
Best Local Similarity 75.0%; Pred. No. 10;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 RPIKPSWAPK 12
      :||| |||
Db     200 KPVPKPSPPAPK 211

```

```

RESULT 10
W4WL18
A:Title: E4 protein - human papillomavirus type 18
C:Species: human papillomavirus type 18
C:Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C:Accession: E26251
R:Cole, S.T.; Danos, O.
J. Mol. Biol. 193, 599-608, 1987
A:Title: Nucleotide sequence and comparative analysis of the human papillomavirus type
A:Reference number: A92937; MUID:87283882; PMID:3039146
A:Accession: E26251
A:Molecule type: DNA
A:Residues: 1-88 <COL>
A:Cross-references: GB:X05015; NID:g60975; PIDN:CAA28668.1; PID:g60980
C:Superfamily: papillomavirus E4 protein
C:Keywords: early protein

Query Match          37.3%; Score 53; DB 1; Length 88;
Best Local Similarity 45.8%; Pred. No. 3.8;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

Qy      3 IPKPSWAPK----KHKRLSDQDS 22
      ||| ||| |||: |||: |||
Db     28 IPAPCPWAPQRPRTARRLLHDLDT 51

RESULT 11
T40820
A:Title: Proline-rich protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: T40820
R:Beck, A.; Reinhardt, R.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z21949
A:Accession: T40820
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-238 <BEC>
A:Cross-references: EMBL:AL032684; PIDN:CAA21811.1; GSPDB:GN00067; SPDB:SPBP8B7.26
A:Experimental source: strain 972h-; clone pl p8B7
C:Genetics:
A:Gene: SPDB:SPBP8B7.26
A:Map position: 2
C:Superfamily: proline-rich protein

Query Match          37.3%; Score 53; DB 2; Length 238;
Best Local Similarity 52.4%; Pred. No. 10;
Matches 11; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy      2 PIPKPSWAPKHKRRRLSDQDS 22
      :||| ||| |||: |||: |||
Db     194 PVKAPPPAPPPKPRLLAPTS 214

RESULT 12
A36264
A:Title: DNA-directed RNA polymerase (EC 2.7.7.6) II 33K chain - human
C:Species: Homo sapiens (man)
C:Date: 18-Jan-1991 #sequence_revision 18-Jan-1991 #text_change 18-Jun-1999
C:Accession: A36264
R:Pati, U.K.; Weissman, S.M.
J. Biol. Chem. 265, 8400-8403, 1990
A:Title: The amino acid sequence of the human RNA polymerase II 33-kDa subunit hRBP 33
A:Reference number: A36264; MUID:90256750; PMID:2187864
A:Accession: A36264
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-275 <PAT>
A:Cross-references: GB:J05448; NID:g337496; PIDN:AAA36586.1; PID:g337497
C:Superfamily: DNA-directed RNA polymerase II chain RPB3; ferredoxin 2[4Fe-4S] homology

```

C;Keywords: nucleotidyltransferase; transcription

Query Match 37.3%; Score 53; DB 2; Length 275;
Best Local Similarity 36.4%; Pred. No. 12;
Matches 8; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 PKPSPWAPKXRRRLSDQDSQTP 25
||| : : : : :
Db 198 KPPEWPKSEYSELDESSQAP 219

RESULT 13

JC4609
chitin synthase (EC 2.4.1.16) A - Ampelomyces quisqualis
N;Alternate names: ChsA protein
C;Species: Ampelomyces quisqualis
C;Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000
C;Accession: JC4609
R;Weiss, N.; Szejnberg, A.; Yarden, O.
Gene 168, 99-102, 1996
A;Title: The chsA gene, encoding a class-I chitin synthase from Ampelomyces quisqualis.
A;Reference number: JC4609; MUID:96186963; PMID:8626074
A;Accession: JC4609
A;Molecule type: DNA
A;Residues: 1-910 <WEI>
A;Cross-references: EMBL:X86802; NID:gl486268; PIDN:CAA60497.1; PID:gl486269
C;Comment: This enzyme belongs to the class-I chitin synthases, and is a member of a mul
C;Genetics:
A;Gene: chsA
A;Introns: 324/3
C;Function:
A;Description: catalyzes the alpha-1,4-glycosylation of chitin by UDP-N-acetyl-D-glucosa
C;Superfamily: chitin synthase chsA
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 37.3%; Score 53; DB 2; Length 910;
Best Local Similarity 60.0%; Pred. No. 39;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 4 PKPSPWAPKXRRRLS 18
| : : : : :
Db 7 PSRSPWAPTQRRRLA 21

RESULT 14

H96646
hypothetical protein F8K4.24 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H96646
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96646
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-684 <STO>
A;Cross-references: GB:AB005173; NID:g3367537; PIDN:AAC28522.1; GSPDB:GN00141
C;Genetics:
A;Gene: F8K4.24
A;Map position: 1

Query Match 37.0%; Score 52.5; DB 2; Length 684;
Best Local Similarity 35.7%; Pred. No. 34;
Matches 10; Conservative 5; Mismatches 8; Indels 5; Gaps 1;

QY 3 IPKPSPW-----APKKHRLSDQDSQTP 25
| : : : : :
Db 502 ISRSTWLRLATGKSHRRVVEESQTP 529

RESULT 15

T02149
hypothetical protein F8K4.24 - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 22-Oct-1999
C;Accession: T02149
R;Vyotskaia, V.S.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu, G.; Oji, O.; Liu, S.; Li,
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel
submitted to the EMBL Data Library, August 1998
A;Description: Arabidopsis thaliana chromosome 1 BAC F8K4 sequence.
A;Reference number: Z14574
A;Accession: T02149
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-684 <VYS>
A;Cross-references: EMBL:AC004392; NID:g3282170; PID:g3367537; GSPDB:GN00059; ATSP:F8K4

C;Genetics:
A;Gene: ATSP:F8K4.24
A;Map position: 1
A;Introns: 443/3; 551/1; 610/3

Query Match 37.0%; Score 52.5; DB 2; Length 684;
Best Local Similarity 35.7%; Pred. No. 34;
Matches 10; Conservative 5; Mismatches 8; Indels 5; Gaps 1;

QY 3 IPKPSPW-----APKKHRLSDQDSQTP 25
| : : : : :
Db 502 ISRSTWLRLATGKSHRRVVEESQTP 529

Search completed: May 27, 2004, 16:21:43
Job time : 18.8269 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:03:58 ; Search time 11.5385 Seconds
(without alignments)
112.819 Million cell updates/sec

Title: US-10-008-524A-4
Perfect score: 142
Sequence: 1 RPIKPSFPWAPKXRRRLSDQSQTP 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	121	85.2	95	VE4_HPV16	P06922 human papil
2	80	56.3	96	VE4_HPV35	P27224 human papil
3	74	52.1	87	VE4_HPV51	P46548 human papil
4	70.5	49.6	102	VE4_HPV31	P17384 human papil
5	61.5	43.3	1615	RIM1_RAT	Q9jir4 rattus norv
6	61.5	43.3	1692	RIM1_HUMAN	Q86ur5 homo sapien
7	59	41.5	94	VE4_HPV39	P24831 human papil
8	54	38.0	670	ZN16_HUMAN	P17020 homo sapien
9	53	37.3	88	VE4_HPV18	P06791 human papil
10	53	37.3	275	RPB3_HUMAN	P19387 homo sapien
11	53	37.3	755	MTS1_HUMAN	Q43312 homo sapien
12	53	37.3	910	CHSA_AMPOU	Q12564 ampelomyces
13	52.5	37.0	1530	RIM2_MOUSE	Q9eqz7 mus musculus
14	52	36.6	459	MIG6_RAT	P05432 rattus norv
15	52	36.6	461	MIG6_MOUSE	Q99jz7 mus musculus
16	52	36.6	1508	GEM5_HUMAN	Q8teq6 homo sapien
17	51	35.9	101	VE25_BPML5	Q52232 mycobacteri
18	51	35.9	759	MTS1_MOUSE	Q8r184 mus musculus
19	51	35.9	1039	ITAB_HUMAN	P08514 homo sapien
20	50.5	35.6	1142	ENAM_HUMAN	Q9nrm1 homo sapien
21	50	35.2	138	H5_COLLI	P02260 columba liv
22	50	35.2	104	HGL5_CHICK	P12902 gallus gall
23	49.5	34.9	361	IHA_FRIVO	Q7755 trichosurus
24	49.5	34.9	1188	RIM2_HUMAN	Q9uq26 homo sapien
25	49.5	34.9	1555	RIM2_MOUSE	Q9jisi rattus norv
26	49	34.5	101	VE25_BPMD2	Q64219 mycobacteri
27	49	34.5	135	INL5_HUMAN	Q9y5q6 homo sapien
28	49	34.5	325	CTH1_YEAST	P47976 saccharomyc
29	49	34.5	393	TRMB_HELPFY	Q25443 helicobacte
30	49	34.5	400	TRMB_HELPF	Q9z196 helicobacte
31	49	34.5	893	YM92_CAEEL	P34531 caenorhabdi
32	49	34.5	956	YEF3_YEAST	P32618 saccharomyc
33	49	34.5	1207	DML1_ARATH	Q9sjg6 arabidopsis

34	49	34.5	1628	1	YATE_SCHPO	Q09779 schizosacch
35	48.5	34.2	1513	1	GRLF_HUMAN	Q9nry4 homo sapien
36	48	33.8	282	1	GDA3_WHEAT	P04723 triticum ae
37	48	33.8	296	1	GDA6_WHEAT	P04726 triticum ae
38	48	33.8	305	1	AMP3_SYNY3	P53581 synechocyst
39	48	33.8	452	1	GAT2_XENLA	P23770 xenopus lae
40	48	33.8	462	1	MIG6_HUMAN	Q9ujm3 homo sapien
41	48	33.8	1032	1	VG07_BPT4	P19061 bacterioph
42	47.5	33.5	535	1	SPKC_SYNY3	P74745 synechocyst
43	47.5	33.5	601	1	CIK5_MUSPF	P79197 mustela put
44	47	33.1	182	1	FTN_DROME	P35554 drosophila
45	47	33.1	210	1	TRMB_MYCGE	P47589 mycoplasma

ALIGNMENTS

RESULT 1				
ID	VE4 HPV16	STANDARD;	PRT;	95 AA.
AC	P06922;			
DT	01-JAN-1988 (Rel. 06, Created)			
DT	01-JAN-1988 (Rel. 06, Last sequence update)			
DT	01-JUL-1993 (Rel. 26, Last annotation update)			
DE	Probable E4 protein.			
OS	Human papillomavirus type 16.			
OC	Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;			
OC	Papillomavirus.			
OX	NCBI_TaxID=10581;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85246220; PubMed=2990099;			
RA	Seedorf K., Krammer G., Durst M., Suhai S., Rowekamp W.G.;			
RT	"Human papillomavirus type 16 DNA sequence.";			
RL	Virology 145:181-185(1985).			
RN	[2]			
RP	SEQUENCE OF 6-95 FROM N.A.			
RX	MEDLINE=90218027; PubMed=2157796;			
RA	Schneider-Maunoury S., Pebau-Arnaudet G., Breitburd F., Orth G.;			
RT	"Expression of the human papillomavirus type 16 genome in SK-v cells,			
RT	a line derived from a vulvar intraepithelial neoplasia.";			
RL	J. Gen. Virol. 71:809-817(1990).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; K02718; AAA46937.1; --			
DR	EMBL; D00735; BAA00634.1; --			
DR	PIR; A22355; W4WLS.			
DR	InterPro; IPR003861; Papilloma_E4.			
DR	Pfam; PF02711; Pap_E4; 1.			
KW	Early protein.			
SQ	SEQUENCE 95 AA; 10594 MW; AED4269D177307CE CRC64;			

Query Match 85.2%; Score 121; DB 1; Length 95;
Best Local Similarity 92.6%; Pred. No. 7.3e-09;
Matches 25; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy	1	RPIKPSFPWAPKXRRRL-SDQD-SQTP 25
Db	29	RPIKPSFPWAPKXRRRLSSDQSQSQTP 55

RESULT 2			
VE4_HPV35			
ID	VE4_HPV35	STANDARD;	PRT;
AC	P27224;		
DT	01-AUG-1992 (Rel. 23, Created)		

```
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 35.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10587;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92124753; PubMed=1310198;
RA Marich J.B., Pontsler A.V., Rice S.M., McGraw K.A., Dubensky T.W.;
RT "The phylogenetic relationship and complete nucleotide sequence of
RL human papillomavirus type 35.";
RL Virology 186:770-776(1992)
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M74117; AAA46965.2; -.
DR PIR; C40824; W4WL35.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.
SQ SEQUENCE 96 AA; 10597 MW; AE4524418CD26F7C CRC64;

Query Match          56.3%; Score 80; DB 1; Length 96;
Best Local Similarity 66.7%; Pred. No. 0.0011;
Matches 14; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 RPIKPSWPAPKKHRRLSQDQ 21
   |||||:||||:||||
DB 29 RPIKPAPWAPQKPRRQITND 49

RESULT 3
VE4_HPV51
ID VE4 HPV51 STANDARD; PRT; 87 AA.
AC P26548;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DE Probable E4 protein.
OS Human papillomavirus type 51.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10595;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91303675; PubMed=1649326;
RA Lungu O., Crum C.P., Silverstein S.J.;
RT "Biologic properties and nucleotide sequence analysis of human
RL papillomavirus type 51.";
RL J. Virol. 65:4216-4225(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M62877; -; NOT ANNOTATED CDS.
DR PIR; C40415; W4WL51.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.
SQ SEQUENCE 87 AA; 9941 MW; 5F3DC38F86BF3990 CRC64;
```

```
Query Match          52.1%; Score 74; DB 1; Length 87;
Best Local Similarity 63.6%; Pred. No. 0.006;
Matches 14; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 RPIKPSWPAPKKHRRLSQDQ 22
   |||||:||||:||||
DB 24 RPIPLPAPWAPKPRHNSENDS 45

RESULT 4
VE4_HPV31
ID VE4 HPV31 STANDARD; PRT; 102 AA.
AC P17384;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE Probable E4 protein.
OS Human papillomavirus type 31.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10585;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89299478; PubMed=2545036;
RA Goldsborough M.D., Disilvestre D., Temple G.F., Lorincz A.T.;
RT "Nucleotide sequence of human papillomavirus type 31: a cervical
RT neoplasia-associated virus.";
RL Virology 171:306-311(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04353; AAA46949.1; -.
DR PIR; E32444; W4WL31.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.
SQ SEQUENCE 102 AA; 11284 MW; 04E3C9E8ABC5CA6C CRC64;

Query Match          49.6%; Score 70.5; DB 1; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.019;
Matches 16; Conservative 4; Mismatches 3; Indels 9; Gaps 2;

QY 3 IPKPSWPAPKK-----HRRLSQDQ----SQTP 25
   |||||:||||:||||:||||:
DB 32 IPKPAPWAPVKVCGRRRLSDQSQSQSTETP 63

RESULT 5
RIM1_RAT
ID RIM1 RAT STANDARD; PRT; 1615 AA.
AC Q9JUE4; Q35168;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Regulating synaptic membrane exocytosis prtein 1 (Rab3-interacting
DE molecule 1) (RIM 1).
GN RIMS1 OR RIM1
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RAB3A AND RAB3C.
RC TISSUE=Brain;
RX MEDLINE=97394473; PubMed=9252191;
RA Wang Y., Okamoto M., Schmitz F., Hofmann K., Suedhof T.C.;
```


Comment=Additional isoforms seem to exist;
 Name=1;
 IsoId=Q86UR5-1; Sequence=Displayed;
 Name=2; Synonyms=RIM short form;
 IsoId=Q86UR5-2; Sequence=VSP_008160, VSP_008165;
 Name=3; Synonyms=RIM long form, Rab3 interacting protein variant 2;
 IsoId=Q86UR5-3; Sequence=VSP_008160, VSP_008164, VSP_008167, VSP_008171;
 Name=4; Synonyms=Rab3 interacting protein variant 1;
 IsoId=Q86UR5-4; Sequence=VSP_008166, VSP_008167;
 Name=5; Synonyms=Rab3 interacting protein variant 3;
 IsoId=Q86UR5-5; Sequence=VSP_008161, VSP_008163, VSP_008170;
 Name=6; Synonyms=Rab3 interacting protein variant 4;
 IsoId=Q86UR5-6; Sequence=VSP_008164, VSP_008168, VSP_008169;
 Name=7; Synonyms=Rab3 interacting protein variant 5;
 IsoId=Q86UR5-7; Sequence=VSP_008161, VSP_008164, VSP_008167, VSP_008169;
 Name=8; Synonyms=Rab3 interacting protein variant 6;
 IsoId=Q86UR5-8; Sequence=VSP_008161, VSP_008162, VSP_008169;
 TISSUE SPECIFICITY: Detected in brain and retina.
 DISEASE: Defects in RIMS1 are a cause of autosomal dominant congenital rod dystrophy (CORD7) [MIM:603649]. CORD7 is characterized by early loss of visual acuity and color vision, followed by night blindness and peripheral visual field loss. The onset of reduced color vision and visual acuity varies between the ages of 20 and 40 years.
 SIMILARITY: Contains 2 C2 domains.
 SIMILARITY: Contains 1 FYVE-type zinc finger.
 SIMILARITY: Contains 1 PDZ/DHR domain.
 SIMILARITY: Contains 1 Rab-binding (RabBD) domain.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 EMBL; AY190519; AAC38848.1; -;
 EMBL; AB045726; BAB87121.1; -;
 EMBL; AB051866; BAB87242.1; -;
 EMBL; AB020338; BAA20798.1; ALT_INIT.
 EMBL; AF263305; AAG23162.1; -;
 EMBL; AF263306; AAG23163.1; -;
 EMBL; AF263307; AAG23164.1; -;
 EMBL; AF263308; AAG23165.1; -;
 EMBL; AF263309; AAG23166.1; -;
 EMBL; AF263310; AAG23167.1; -;
 HSSP; P21707; 1RSY.
 Genew; HGNC:17282; RIMS1.
 MIM; 606629; -;
 MIM; 603649; -;
 InterPro; IPR000008; C2.
 InterPro; IPR008973; C2 CaLB.
 InterPro; IPR001478; PDZ.
 InterPro; IPR003315; RPH3A effector.
 InterPro; IPR000306; Znf_FYVE.
 Pfam; PF00168; C2; 2.
 Pfam; PF00595; PDZ; 1.
 Pfam; PF02318; RPH3A effector; 1.
 SMART; SM00239; C2; 2.
 SMART; SM00228; PDZ; 1.
 PROSITE; PS50004; C2_DOMAIN_2; 2.
 PROSITE; PS50106; PDZ; 1.
 PROSITE; PS50916; RABBD; 1.
 PROSITE; PS50178; ZF_FYVE; 1.
 Metal-binding; Zinc; Repeat; Zinc-finger; Alternative splicing;
 Vision; Disease mutation.
 DOMAIN 22 182 RAB-BINDING.
 FT DOMAIN 110 170 FYVE-TYPE.
 FT ZN_FING 605 691 PDZ.
 FT DOMAIN

FT	DOMAIN	744	850	C2 DOMAIN 1.
FT	DOMAIN	1538	1640	C2 DOMAIN 2.
FT	DOMAIN	1336	1402	SER-RICH.
FT	VARSPLIC	484	494	Missing (in isoform 2 and isoform 3).
FT	VARSPLIC	924	924	Missing (in isoform 5, isoform 7 and isoform 8).
FT	VARSPLIC	1018	1245	Missing (in isoform 8).
FT	VARSPLIC	1038	1244	Missing (in isoform 5).
FT	VARSPLIC	1039	1102	Missing (in isoform 3, isoform 6 and isoform 7).
FT	VARSPLIC	1040	1692	Missing (in isoform 2).
FT	VARSPLIC	1065	1102	Missing (in isoform 4).
FT	VARSPLIC	1133	1245	Missing (in isoform 3, isoform 4 and isoform 7).
FT	VARSPLIC	1161	1245	Missing (in isoform 6).
FT	VARSPLIC	1284	1455	Missing (in isoform 6, isoform 7 and isoform 8).
FT	VARSPLIC	1377	1385	Missing (in isoform 5).
FT	VARSPLIC	1540	1573	Missing (in isoform 3).
FT	VARIANT	820	820	R -> H (in CORD7).
FT	MUTAGEN	796	797	FTId=VAR_016804.
FT	MUTAGEN	1591	1592	RR->AA: ABOLISHES INTERACTION WITH SYT1 AND CACNA1B.
FT	SEQUENCE	1692 AA; 189071 MW; 0A96642DC832C15E CRC64;		KK->AA: ABOLISHES INTERACTION WITH SYT1 AND CACNA1B.
Query Match		43.3%;	Score 61.5;	DB 1; Length 1692;
Best Local Similarity		33.3%;	Pred. No. 4;	
Matches	11; Conservative	7; Mismatches	2; Indels	13; Gaps
Qy	2 PIPKSPWAPKKH-----RLSDQD 21			
Db	876 PLQPSPPMPRRHIGSSSKLQSRQISDS 908			
RESULT 7				
VE4 HPV39				
ID VE4 HPV39	STANDARD;	PRT;	94 AA.	
AC P24831;				
DT 01-MAR-1992 (Rel. 21, Created)				
DT 01-MAR-1992 (Rel. 21, Last sequence update)				
DT 01-MAR-1992 (Rel. 21, Last annotation update)				
DE Probable E4 protein.				
OS Human papillomavirus type 39.				
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;				
OC Papillomavirus.				
OX NCBI_taxid=10588;				
RN [1]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=91135017; PubMed=1847266;				
RA Volpers C., Streack R.E.;				
RT "Genome organization and nucleotide sequence of human papillomavirus type 39.";				
RL Virology 181:419-423 (1991).				
CC -----				
CC This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC the European Bioinformatics Institute. There are no restrictions on its				
CC use by non-profit institutions as long as its content is in no way				
CC modified and this statement is not removed. Usage by and for commercial				
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC or send an email to license@isb-sib.ch).				

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

EMBL; X05015; CAA28668.1; -
DR PIR; E26251; W4WU18.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.
SQ SEQUENCE 88 AA; 9857 MW; D05F6200BF367B10 CRC64;

Query Match 37.3%; Score 53; DB 1; Length 88;
Best Local Similarity 45.8%; Pred. No. 2.8;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

QY 3 IPKPSWPAPX---KHRELSDQS 22
||| ||||| : ||||| :
Db 28 IPAPCPWAPQRPRTARRRLHDLDT 51

RESULT 10

RPB3 HUMAN
ID_RPB3 HUMAN STANDARD; PRT; 275 AA.
AC P19387; O15161;
DT 01-NOV-1990 (Rel. 16, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-directed RNA polymerase II 33 kDa polypeptide (EC 2.7.7.6) (RPB3)
DE (RNA polymerase II subunit 3) (RPB33) (RPB31).
GN POLR2C OR A-152E5.7.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_taxid:9606;
EN [1]

RP SEQUENCE FROM N.A. PubMed=2187864;
RX MEDLINE=90256750; PubMed=2187864;
RA Pati U.K., Weissman S.M.;
RT "The amino acid sequence of the human RNA polymerase II 33-kDa
RT subunit hRPB 33 is highly conserved among eukaryotes.";
RL J. Biol. Chem. 265:8400-8403(1990).
RN [2]

RP SEQUENCE FROM N.A. PubMed=9540830;
RX MEDLINE=98201707; PubMed=9540830;
RA Damann R., Pfeifer G.P.;
RT "Cloning and characterization of the human RNA polymerase I subunit
RT hRPA40.";
ET hRPA40.";
RL Biochim. Biophys. Acta 1396:153-157(1998).
RN [3]

RP SEQUENCE FROM N.A.
RX TISSUE=Skeletal muscle;
RA Bruno T., di Padova M., de Angelis R., Iacobini C., Lovari S.,
RA Passananti C., Fanciulli M.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [4]

RP SEQUENCE FROM N.A.
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Deslattes Mays A., Cao X., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
RT human chromosome 16p and 16q.";
RL Genomics 60:295-308(1999).
RN [5]

RP SEQUENCE FROM N.A.
RX TISSUE=Brain, Kidney, and Muscle;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC POLR2J AND POLR2C SUBUNITS INTERACT WITH EACH OTHER.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are
CC found in eukaryotic nuclei: polymerase I for the ribosomal RNA
CC precursor, polymerase II for the mRNA precursor, and polymerase
CC III for 5S and tRNA genes.
CC -!- SIMILARITY: Belongs to the archaeobacteria RPOD / eukaryotic RPB3
CC RNA polymerase subunit family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed.
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J05448; AAA36586.1; -
DR EMBL; AF008443; AAC14355.1; -
DR EMBL; AJ224143; CAAL1842.1; -
DR EMBL; AJ224144; CAAL1843.1; -
DR EMBL; AC004382; AAC24309.1; -
DR EMBL; BC000409; AAH00409.1; -
DR EMBL; BC003159; AAH03159.1; -
DR EMBL; BC028157; AAH28157.1; -
DR PIR; A36264; A36264.
DR Genew; HGNC:9189; POLR2C.
DR MIM; 180663; -
DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.
DR InterPro; IPR009025; RBP11-like RNAPo.
DR InterPro; IPR001700; RNA_POLA_bac_org.
DR InterPro; IPR001514; RNA_POLD.
DR Pfam; PF01000; RNA_pol_A_bac; 1.
DR ProDom; PD028883; RNA_POLD; 1.
DR SMART; SM00662; RPOLD; 1.
DR PROSITE; PS00446; RNA_POL_D_30KD; 1.
DR Transfaser; DNA-directed RNA polymerase; Transcription;
KW Nuclear protein.
FT DOMAIN 88 97 CYS-RICH.
FT CONFLICT 194 194 H -> T (IN REF. 1 AND 3).
SQ SEQUENCE 275 AA; 31441 MW; EF63BE096046A4B CRC64;

Query Match 37.3%; Score 53; DB 1; Length 275;
Best Local Similarity 36.4%; Pred. No. 8.3;
Matches 8; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 PKPSWPAPKXRRRLSDQSQTP 25
||||| : : : : :
Db 198 PKPEWPAPKSEYSELDEDESQAP 219

RESULT 11

Biochem. J. 371:463-471(2003).
 -!- FUNCTION: May be related to cancer progression or tumor metastasis in a variety of organ sites, most likely through an interaction with the actin cytoskeleton.
 -!- SUBUNIT: Binds to actin. Binds to the cytoplasmic domain of receptor protein tyrosine phosphatase delta.
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=3;
 Name=1;
 IsoId=O43312-1; Sequence=Displayed;
 Name=2;
 IsoId=O43312-2; Sequence=VSP_007420, VSP_007421;
 Note=No experimental confirmation available;
 Name=3;
 IsoId=O43312-3; Sequence=VSP_007419;
 -!- TISSUE SPECIFICITY: Expressed in many tissues, including spleen, thymus, prostate, testis, uterus, colon, and peripheral blood.
 -!- DOMAIN: The WH2 motif at the C-terminus binds to actin monomers.
 -!- SIMILARITY: Contains 1 WH2 domain.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; AF086645; AAF15947.1; --
 EMBL; AK027015; --; NOT_ANNOTATED_CDS.
 EMBL; AB007889; BAA24859.1; ALT_INIT.
 EMBL; BC023998; AAH23998.1; --
 Genew; HGNC:20443; MTSS1.
 InterPro: IPR003124; WH2.
 Pfam; PF02205; WH2; 1.
 DR Cytoskeleton; Actin-binding; Coiled coil; Anti-oncogene;
 KW Alternative splicing.
 FT DOMAIN 727 744 WH2.
 FT FT DOMAIN 108 155 COILED COIL (POTENTIAL).
 FT FT DOMAIN 238 359 SER-RICH.
 FT FT DOMAIN 608 726 PRO-RICH.
 FT FT VARSPPLIC 1 344 Missing (in isoform 3).
 FT FT VARSPPLIC 1 200 Missing (in isoform 2).
 FT FT VARSPPLIC 345 426 Missing (in isoform 2).
 FT FT CONFLICT 153 153 K -> KVDTL (IN REF. 2).
 FT FT CONFLICT 586 586 V -> A (IN REF. 2).
 FT FT CONFLICT 630 630 L -> M (IN REF. 1).
 SQ SEQUENCE 755 AA; 82250 MW; 76PEE3224CBA9287 CRC64;
 Query Match 37.3%; Score 53; DB 1; Length 755;
 Best Local Similarity 33.3%; Pred.No.22;
 Matches 10; Conservative 6; Mismatches 8; Indels 6; Gaps 1;
 QY 2 PYPKSPWAPKKHR-----LSDQDSQTP 25
 DB 674 PUPGPKPSIEBHRQAIPESEAEDQEREPP 703
 RESULT 12
 CHSA_AMPOU
 ID CHSA_AMPOU STANDARD; PRT; 910 AA.
 AC Q12564;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chitin synthase A (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl
 DE transferase A) (Class-I chitin synthase A).
 GN CHSA.
 OS Ampelomyces quisqualis.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;

CC Pleosporales; Leptosphaeriaceae; mitosporic Leptosphaeriaceae;
 CC Ampelomyces.
 CC NCBI TaxID=50730;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AQ10;
 RX MEDLINE=96186963; PubMed=8626074;
 RA Weiss N., Szejnberg A., Yarden O.;
 RT "The chsA gene, encoding a class-I chitin synthase from *Ampelomyces*
 RT *quisqualis*.";
 RL Gene 168:99-102(1996).
 CC -!- FUNCTION: Plays a major role in cell wall biogenesis.
 CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + {(1,4)-(N-acetyl-
 CC beta-D-glucosaminyl)}(N) = UDP + {(1,4)-(N-acetyl-beta-D-
 CC glucosaminyl)}(N+1).
 CC -!- SUBCELLULAR LOCATION: Plasma membrane-bound.
 CC -!- SIMILARITY: Belongs to the chitin synthase family. Subfamily class
 CC I.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X86802; CAA60497.1; -;
 DR PIR; J04609; JC4609.
 DR InterPro; IPR004834; Chitin synth.
 DR InterPro; IPR001173; Glyco_trans_2.
 DR Pfam; PF01644; Chitin synth; 1.
 DR ProDom; PD002998; Chitin synth; 1.
 KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;
 KW Multigene family.
 FT TRANSMEM 366 386 POTENTIAL.
 FT TRANSMEM 448 468 POTENTIAL.
 FT TRANSMEM 583 603 POTENTIAL.
 FT TRANSMEM 620 640 POTENTIAL.
 FT TRANSMEM 655 675 POTENTIAL.
 FT TRANSMEM 701 721 POTENTIAL.
 FT TRANSMEM 730 750 POTENTIAL.
 FT TRANSMEM 828 848 POTENTIAL.
 FT TRANSMEM 876 896 POTENTIAL.
 SQ SEQUENCE 910 AA; 103012 MW; F8F1DB135F2138F9 CRC64;
 Query Match 37.3%; Score 53; DB 1; Length 910;
 Best Local Similarity 60.0%; Pred. No. 26;
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 4 PKPSFWAPKRRRLS 18
 | | | | | : | | :
 Db 7 PSRSPWAPQRRRLA 21
 RESULT 13
 ID RIM2 MOUSE
 AC Q9EQZ7; Q8C433; STANDARD; PRT; 1530 AA.
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Regulating synaptic membrane exocytosis protein 2 (Rab3-interacting
 DE molecule 2) (RIM 2) (Rab3 interacting protein 2).
 GN RIMS2 OR RIM2 OR RAB3IP2.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), TISSUE SPECIFICITY, AND INTERACTION
 CC WITH RAB3A AND EPAC2.
 CC TISSUE=Insulinoma;

RE MEDLINE=20512528; PubMed=11056535;
 RA Ozaki N., Shibasaki T., Kashima Y., Miki T., Takahashi K., Ueno H.,
 RA Sunaga Y., Yano H., Matsura Y., Iwanaga T., Takai Y., Seino S.;
 RT "cAMP-GEFII is a direct target of cAMP in regulated exocytosis.";
 RL Nat. Cell Biol. 2:805-811(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC STRAIN=C57BL/6J; TISSUE=Hippocampus, and Olfactory bulb;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nitaido I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Betsel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Haya A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Yasunishi A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
 RA Birney E., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [3]
 RP INTERACTION WITH PCLO.
 RX MEDLINE=22384373; PubMed=12401793;
 RA Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,
 RA Sasaki T., Tajima N., Iwanaga T., Seino S.;
 RT "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of
 RT cAMP-GEFII.Rim2.piccolo complex in cAMP-dependent exocytosis.";
 RL J. Biol. Chem. 277:50497-50502(2002).
 CC -!- FUNCTION: Rab effector involved in exocytosis. May act as scaffold
 CC protein.
 CC -!- SUBUNIT: Binds RAB3A and RAB3B that have been activated by GTP-
 CC binding (by similarity). Heterodimer with PCLO. Part of a ternary
 CC complex involving PCLO and EPAC2.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=Q9EQZ7-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9EQZ7-2; Sequence=VSP_008181, VSP_008184;
 CC Note=No experimental confirmation available;
 CC Name=3;
 CC IsoId=Q9EQZ7-3; Sequence=VSP_008182, VSP_008183;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Detected in testis, pituitary and an
 CC insulinoma cell line. Detected at low levels in cerebellar cortex.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
 CC -!- SIMILARITY: Contains 1 PDZ/DRH domain.
 CC -!- SIMILARITY: Contains 1 Rab-binding (RABBD) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

RA Whitting M., Maud A., Young A.C., Shevchenko I., Bourlard
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.I., Skalska S., Smalusz D.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra MA.

RA Whitting M., Maud A., Young A.C., Shevchenko I., Bourlard
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywicki M.I., Skalska S., Smalusz D.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra MA.

```

RT "generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:16899-16903(2002).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the MIG6 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BC005546; AAH05546.1; -.
DR EMBL; BC057646; AAH57646.1; -.
DR MGD; MGI:1921405; L300002F13Rik.
SQ SEQUENCE 461 AA; 50075 MW; 0CA414516FAE27A6 CRC64;

Query Match 36.6%; Score 52; DB 1; Length 461;
Best Local Similarity 52.9%; Pred.No. 18;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 RPIPKSPWAPKKHRL 17
      | : | | | | |
Db 230 RVVEDNPDPPOSHRL 246

```

Search completed: May 27, 2004, 16:19:02
Job time : 12.5385 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:12:58 ; Search time 48.0769 Seconds
(without alignments)
164.069 Million cell updates/sec

Title: US-10-008-524A-4
Perfect score: 142
Sequence: 1 RPIPKPSFWAPKKHRLSDQDSQTP 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25; *
1: sp archaea: *
2: sp bacteria: *
3: sp fungi: *
4: sp human: *
5: sp invertebrate: *
6: sp mammal: *
7: sp mhc: *
8: sp organelle: *
9: sp phage: *
10: sp plant: *
11: sp rodent: *
12: sp virus: *
13: sp vertebrate: *
14: sp unclassified: *
15: sp virus: *
16: sp bacteriap: *
17: sp archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	121	85.2	95	12	Q918T2 human papill
2	121	85.2	95	12	Q918U0 human papil
3	121	85.2	95	12	Q918T7 human papil
4	121	85.2	95	12	Q8B5P2 human papil
5	121	85.2	95	12	Q8BSN7 human papil
6	121	85.2	95	12	Q8BMM0 human papil
7	72.5	51.1	88	12	Q993Z6 human papil
8	67	47.2	467	10	Q9AX37 oryza sativ
9	67	47.2	467	10	Q7XC25 oryza sativ
10	64	45.1	97	12	Q90727 human papil
11	60	42.3	1242	12	Q9PZW6 eastern equ
12	59	41.5	356	10	Q9L1P1 arabidopsis
13	58	40.8	985	5	Q9Y142 drosophila
14	58	40.8	985	5	Q9V4X8 drosophila
15	57.5	40.5	88	12	Q9IR55 human papil
16	57	40.1	106	12	Q82001 human papil

17	57	40.1	230	16	Q49630 mycobacteri
18	57	40.1	1242	12	Q9PZW8 eastern equ
19	57	40.1	1242	12	Q9PZW7 eastern equ
20	56.5	39.8	102	12	Q8JN97 human papil
21	56	39.4	117	12	Q8JTB9 human papil
22	56	39.4	137	5	Q22204 caenorhabdi
23	56	39.4	1095	10	Q9SM80 oryza sativ
24	54	38.0	330	10	Q49870 hordeum vul
25	54	38.0	671	4	Q9NRA4 homo sapien
26	54	38.0	688	4	Q8ND22 homo sapien
27	54	38.0	1038	5	Q8MQW9 drosophila
28	54	38.0	2347	5	Q8INB9 drosophila
29	54	38.0	2451	5	Q9VG05 drosophila
30	53	37.3	262	3	Q94274 schizosacch
31	53	37.3	275	11	Q9DB22 mus musculu
32	53	37.3	275	11	Q99M46 mus musculu
33	53	37.3	382	10	Q9SHZ5 arabidopsis
34	53	37.3	382	10	Q8L9T4 arabidopsis
35	53	37.3	2527	5	Q95W83 plasmodium
36	52.5	37.0	461	13	Q7ZUL1 brachydanio
37	52.5	37.0	624	10	Q94CF4 arabidopsis
38	52.5	37.0	684	10	Q80709 arabidopsis
39	52	36.6	184	4	Q9BU61 homo sapien
40	52	36.6	234	10	Q9FP79 oryza sativ
41	52	36.6	275	5	Q97183 drosophila
42	52	36.6	316	10	Q7XK97 oryza sativ
43	52	36.6	318	16	Q82MD4 streptomyce
44	51	35.9	79	4	Q8NEF7 homo sapien
45	51	35.9	133	12	Q8B5X3 human papil

ALIGNMENTS

RESULT 1

Q918T2 PRELIMINARY; PRT; 95 AA.
AC Q918T2; 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE E4 protein (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPV16E2CC7;
RA Watts K.J., Thompson C.H., Cossart Y.E., Rose B.R.;
RT "Sequence variation and physical state of human Papillomavirus type 16 cervical cancer isolates from Australia and New Caledonia."
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF407220; AAL01405.1;
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
DR NON TER 1
SQ SEQUENCE 95 AA; 10542 MW; AED17903867307CE CRC64;

Query Match 85.2%; Score 121; DB 12; Length 95;
Best Local Similarity 92.6%; Pred. No. 6.7e-10;
Matches 25; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 RPIPKPSFWAPKKHRL-SDQDSQTP 25

Db 29 RPIPKPSFWAPKKHRLSDQDSQTP 55

RESULT 2

Q918U0 PRELIMINARY; PRT; 95 AA.
ID Q918U0
AC Q918U0;


```

OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Asian-American variant;
RA Terai M., Burk R.D.;
RT "Human papillomavirus type 16 Asian-American variant.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF402678; AAC05412.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON_TER
SQ SEQUENCE 95 AA; 10542 MW; AED17903867307CE CRC64;

Query Match 85.2%; Score 121; DB 12; Length 95;
Best Local Similarity 92.6%; Pred. No. 6.7e-10;
Matches 25; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 RPIPKSPWAPKHHRL-SDOD-SQTP 25
Db 29 RPIPKSPWAPKHHRLSSDQDSQTP 55

RESULT 7
Q99326 ID Q99326 PRELIMINARY; PRT; 88 AA.
AC Q99326;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Putative E4 protein.
GN E4.
OS Human papillomavirus type 82.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=129724;
RN [1]
RP SEQUENCE FROM N.A.
RA Terai M., Burk R.D.;
RT "Cervical HPVs in Evolution; Genomic Sequence of IS39/AEZ, a Subtype
of Oncogenic HPV 82 (WI3B).";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF293961; AAK28453.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
SQ SEQUENCE 88 AA; 10084 MW; 6752D8CF3A9475D7 CRC64;

Query Match 51.1%; Score 72.5; DB 12; Length 88;
Best Local Similarity 50.0%; Pred. No. 0.0061;
Matches 15; Conservative 4; Mismatches 6; Indels 5; Gaps 1;

Qy 1 RPIPKSPWAPKHHRLSDQD-----SQTP 25
Db 24 RPIPKSPWAPKHHRLSSDQDSQTP 53

RESULT 8
Q9AY37 ID Q9AY37 PRELIMINARY; PRT; 467 AA.
AC Q9AY37;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative phosphatidylinositol-4-phosphate 5-kinase.
GN OSJNBA0027P10.22.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;

```

```

RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Hsiao J.,
RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E., Craven B.,
RA Khalak H., Feldblyum T.V., Quackenbush J., White O., Salzberg S.L.,
RA Fraser C.M.;
RT "Oryza sativa chromosome 10 BAC OSJNBA0027P10 genomic sequence.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084763; AAG60194.1; -.
DR Gramene; Q9AY37; -.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR003409; MORN.
DR Pfam; PF02493; MORN; 7.
DR SMART; SMO0698; MORN; 7.
DR KINASE.
SQ SEQUENCE 467 AA; 50906 MW; 8FB85BCC9B980D73 CRC64;

Query Match 47.2%; Score 67; DB 10; Length 467;
Best Local Similarity 68.8%; Pred. No. 0.21;
Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RPIPKSPWAPKHHRR 16
Db 72 RPSPSPPFAPSRRR 87

RESULT 9
Q7XC25 ID Q7XC25 PRELIMINARY; PRT; 467 AA.
AC Q7XC25;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative phosphatidylinositol-4-phosphate 5-kinase.
GN OSJNBA0027P10.22.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017120; AAP55050.1; -.
KW Kinase.
SQ SEQUENCE 467 AA; 50906 MW; 8FB85BCC9B980D73 CRC64;

Query Match 47.2%; Score 67; DB 10; Length 467;
Best Local Similarity 68.8%; Pred. No. 0.21;
Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RPIPKSPWAPKHHRR 16
Db 72 RPSPSPPFAPSRRR 87

RESULT 10
Q90727 ID Q90727 PRELIMINARY; PRT; 97 AA.
AC Q90727;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF E4.
OS Human papillomavirus type 67.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;

```


RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
 RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
 RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,
 RA Park S., Sequeira A., Sethi H., Snir E., Swirskas R.R., Weinburg T.,
 RA Celniker S.E.;
 RT "Full length Drosophila melanogaster cDNA sequence.";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF145629; AA038604.1; -;
 DR FlyBase; FBgn0027585; BcDNA:GH05582.
 SQ SEQUENCE 985 AA; 112852 MW; 6687661C0506BF4F CRC64;

Query Match 40.8%; Score 58; DB 5; Length 985;
 Best Local Similarity 52.4%; Pred. No. 9.2;
 Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 1 RP1PKPSFPWAPKHKRRRLSDQD 21
 || ||||| |||||
 Db 106 RPSHSPSPWQQLSPRNLSD 126

RESULT 14

Q9V4X8 PRELIMINARY; PRT; 985 AA.

AC Q9V4X8;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CG8740 protein.
 GN BcDNA:GH05582 OR CG8740.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;

RA STRAIN=Berkley;
 RC SEQUENCE FROM N.A.
 RX MEDLINE=2019606; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Burtis K.C., Busam D.A., Butler H., Brokstein P., Brottier P.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kraft S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Dou P.L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradscky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G.F., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003836; AAM68836.1; -;
 DR FlyBase; FBgn0027585; BcDNA:GH05582.
 SQ SEQUENCE 985 AA; 112894 MW; 245F18B903DAC636 CRC64;

Query Match 40.8%; Score 58; DB 5; Length 985;
 Best Local Similarity 52.4%; Pred. No. 9.2;
 Matches 11; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 1 RP1PKPSFPWAPKHKRRRLSDQD 21
 || ||||| |||||
 Db 106 RPSHSPSPWQQLSPRNLSD 126

RESULT 15

Q9IR55 PRELIMINARY; PRT; 88 AA.

AC Q9IR55;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE DNA, complete genome.
 GN E4.
 OS Human papillomavirus type 82.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_TaxID=129724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sata T., Matsukura T.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20087389; PubMed=10618284;
 RA Kino N., Sata T., Sato Y., Sugase M., Matsukura T.;

Search completed: May 27, 2004, 16:20:55
Job time : 50.0769 secs

A new method has been developed for detecting a papilloma virus infection in an organism. The method comprises: (i) obtaining a sample of cells from the potential infection site; (ii) contacting the cells with a molecule binding specifically to papilloma virus E4 protein, and (iii) monitoring the binding. The method is useful to detect papilloma virus infections in organisms (especially mammals) and especially HPV infections (e.g. with HPV16, 18, 33, 35, 45, 51, 56, 58 or 61) in humans. Papilloma viruses cause epithelial tumours in humans varying in severity depending on the infection site and HPV type involved. The method is particularly useful to determine papilloma infection in the mammalian cervix and especially to screen for pre-cancerous cervical lesions in

CC humans, since over 90% of cervical carcinoma patients show cervical HPV
 CC infection. It is also useful to determine the type(s) of HPV infection in
 CC a patient, by using a molecule binding specifically to a subset of HPV E4
 CC proteins. This is important, since progression to malignant disease (and
 CC hence clinical prognosis) is dependent on HPV type. Molecules capable of
 CC binding E4 are also useful to target anticancer/antiviral agents capable
 CC of destroying papilloma viruses and/or papilloma virus-infected cells.
 CC The present sequence represents a specifically claimed HPV16 peptide
 CC found in the hydrophilic region

XX Sequence 26 AA;

Query Match 100.0%; Score 91; DB 2; Length 26;

Best Local Similarity 100.0%; Pred. No. 2.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKHR 15
 Db 1 RPIPKPSWAPKKHR 15

RESULT 2

AAU75260
 ID AAU75260 standard; peptide; 26 AA.

XX AC AAU75260;

XX DT 21-MAY-2002 (first entry)

XX DE Hydrophilic region of HPV16 E4 protein.

XX KW Papilloma virus associated antigen; cell proliferation marker;
 KW cervical malignancy; human papillomavirus infection; HPV; lesion;
 KW cellular abnormality; cellular proliferation; cellular growth; dysplasia;
 KW neoplasia; cancer; papilloma smear test; E4 protein.

XX OS Human papillomavirus type 16.

XX PN WO200208764-A1.

XX PD 31-JAN-2002.

XX PF 16-MAR-2001; 2001WO-GB001176.

XX PR 24-JUL-2000; 2000GB-00018140.

XX PA (MEDI-) MEDICAL RES COUNCIL.

XX PI Doorbar J;

XX DR WPI; 2002-188648/24.

XX PS Detecting abnormalities e.g. abnormal cellular proliferation, in a sample
 CC from a patient comprises contacting cells with a molecule which can bind
 CC to a papilloma virus associated antigen, or a cell proliferation or viral
 CC activity marker.

XX PS Disclosure; Page 23; 90pp; English.

XX CC The present invention relates to a method for detecting abnormalities in
 CC a sample from a patient. The method comprises contacting a sample of the
 CC patient's cells with two or more molecules, where at least one molecule
 CC is capable of binding a papilloma virus associated antigen, and at least
 CC one molecule is capable of binding a cell proliferation marker. The
 CC method is useful for simultaneously screening for abnormalities which
 CC indicate or can lead to cervical malignancy, for human papillomavirus
 CC (HPV) infections, and precursor lesions or other conditions which occur
 CC with cervical malignancy. The method is also useful for assessing the
 CC risk associated with cellular abnormality in a patient sample, and for
 CC determining, assessing or diagnosing the presence or absence of abnormal
 CC cellular proliferation, cellular growth abnormality, dysplasia,
 CC neoplasia, or a pre-cancerous or cancerous state in a tissue. The new
 CC method is much simpler, and yields more information more quickly than

CC conventional papilloma smear testing programmes. Compared with previous
 CC methods of screening, the new method has reduced chances of false
 CC negatives occurring, requires fewer samples to gain the same amount of
 CC information, and alleviates the need for repeated or further testing. The
 CC present sequence representing the hydrophilic region of the HPV16 E4
 CC protein binds antibody molecules

XX Sequence 26 AA;

Query Match 100.0%; Score 91; DB 5; Length 26;

Best Local Similarity 100.0%; Pred. No. 2.2e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKHR 15
 Db 1 RPIPKPSWAPKKHR 15

RESULT 3

AAU15564

XX ID AAU15564 standard; protein; 30 AA.

XX AC AAU15564;

XX DT 02-MAR-1992 (first entry)

XX DE Immunoepitope #4 derived from HPV16 E4 peptide.

XX KW cervical cancer; cervical intraepithelial neoplasia; CIN; wart;
 KW squamous cell carcinoma; ELISA; HPV 16.

XX OS Synthetic.

XX PN WO9118294-A.

XX PD 28-NOV-1991.

XX PF 11-MAY-1990; 90SE-00001705.

XX PR 11-MAY-1990; 90SE-00001705.

XX PA (MEDS-) MEDSCAND AB.

XX PI Dillner J, Dillner L, Cheng HM;

XX DR WPI; 1991-369390/50.

XX PT Diagnosis of human papilloma virus infection and PV-carrying tumours -
 CC using synthetic peptide(s) to detect virus specific antigen-antibody
 CC complexes by immunoassay.

XX PS Disclosure; Page 38; 72pp; English.

XX CC This is one of a large number of peptides which have been synthesised on
 CC the basis of the amino acid sequences for the E2, E4, E7, L1 or L2
 CC proteins of HPV 1, 5, 6, 8, 11, 16, 18, 31 and 33. The selection of
 CC peptide sequences was based on the assumption that an immunoreactive
 CC region might be situated in the same relative region of a protein from
 CC different HPV types. The peptides were used in diagnostic immunoassays to
 CC detect HPV-infection. See AAU15523-R15601

XX Sequence 30 AA;

Query Match 100.0%; Score 91; DB 2; Length 30;

Best Local Similarity 100.0%; Pred. No. 2.6e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKHR 15
 Db 7 RPIPKPSWAPKKHR 21

RESULT 4


```

AAW62278
ID AAW62278 standard; peptide; 15 AA.
XX
AC AAW62278;
XX
XX 24-SEP-1998 (first entry)
XX
DE HPV16 hydrophilic region peptide #2.
XX
KW HPV16; human papilloma virus; epithelial tumour; cervical cancer;
KW precancerous cervical lesion; screening; infection; cervix;
KW HPV E4.
XX
OS Human papillomavirus.
XX
XX WO9825145-A1.
XX
XX 11-JUN-1998.
XX
XX 03-DEC-1997; 97WO-GB003321.
XX
XX 03-DEC-1996; 96GB-00025142.
XX
XX 05-SEP-1997; 97GB-00018745.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
XX Doorbar J;
XX
XX WPI; 1998-333497/29.
XX
XX Detecting papilloma virus infection using molecule binding to E4 protein
XX - useful, e.g. in screening for pre-cancerous cervical lesions and to
XX determine type(s) of human papilloma virus infecting human patients.
XX
XX Claim 10; Page 37; 52pp; English.
XX
XX A new method has been developed for detecting a papilloma virus infection
XX in an organism. The method comprises: (i) obtaining a sample of cells
XX from the potential infection site; (ii) contacting the cells with a
XX molecule binding specifically to papilloma virus E4 protein, and (iii)
XX monitoring the binding. The method is useful to detect papilloma virus
XX infections in organisms (especially mammals) and especially HPV
XX infections (e.g. with HPV16, 18, 33, 35, 45, 51, 56, 58 or 61) in humans.
XX Papilloma viruses cause epithelial tumours in humans varying in severity
XX depending on the infection site and HPV type involved. The method is
XX particularly useful to determine papilloma infection in the mammalian
XX cervix and especially to screen for pre-cancerous cervical lesions in
XX humans, since over 90% of cervical carcinoma patients show cervical HPV
XX infection. It is also useful to determine the type(s) of HPV infection in
XX a patient, by using a molecule binding specifically to a subset of HPV E4
XX proteins. This is important, since progression to malignant disease (and
XX hence clinical prognosis) is dependent on HPV type. Molecules capable of
XX binding E4 are also useful to target anticancer/antiviral agents capable
XX of destroying papilloma viruses and/or papilloma virus-infected cells.
XX The present sequence represents a specifically claimed HPV16 peptide
XX found in the hydrophilic region
XX
SQ Sequence 15 AA;
Query Match 90.1%; Score 82; DB 2; Length 15;
Best Local Similarity 93.3%; Pred. No. 0.0002;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 RPIKPSWPAPKKHR 15
| | | | | | | | | |
Db 1 RRIKPSWPAPKKHR 15
| | | | | | | | | |

RESULT 5
AAAR15562
ID AAR15562 standard; protein; 20 AA.
XX
XX AAR15562;
AC AAR15562;

```

```

XX
DT 02-MAR-1992 (first entry)
XX
DE Immunopeptide #2 derived from HPV16 E4 peptide.
XX
XX cervical cancer; cervical intraepithelial neoplasia; CIN; wart;
XX squamous cell carcinoma; ELISA; HPV 16.
XX
OS Synthetic.
XX
XX WO9118294-A.
XX
XX 28-NOV-1991.
XX
XX 11-MAY-1990; 90SE-00001705.
XX
XX 11-MAY-1990; 90SE-00001705.
XX
PA (MEDS-) MEDSCAND AB.
XX
XX Dillner J, Dillner L, Cheng HM;
XX
XX WPI; 1991-369390/50.
XX
XX Diagnosis of human papilloma virus infection and pv-carrying tumours -
XX using synthetic peptide(s) to detect virus specific antigen-antibody
XX complexes by immunoassay.
XX
XX Disclosure; Page 38; 72pp; English.
XX
XX This is one of a large number of peptides which have been synthesised on
XX the basis of the amino acid sequences for the E2, E4, E7, L1 or L2
XX proteins of HPV 1, 5, 6, 8, 11, 16, 18, 31 and 33. The selection of
XX peptide sequences was based on the assumption that an immunoreactive
XX region might be situated in the same relative region of a protein from
XX different HPV types. The peptides were used in diagnostic immunoassays to
XX detect HPV-infection. See AAR15523-R15601
XX
SQ Sequence 20 AA;
Query Match 86.8%; Score 79; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00068;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 IPKPSPWAPKKHR 15
| | | | | | | | | |
Db 1 IPKPSPWAPKKHR 13
| | | | | | | | | |

RESULT 6
AAW62279
ID AAW62279 standard; peptide; 12 AA.
XX
XX AAW62279;
AC AAW62279;
XX
XX 24-SEP-1998 (first entry)
XX
XX HPV16 hydrophilic region peptide #3.
XX
XX HPV16; human papilloma virus; epithelial tumour; cervical cancer;
XX precancerous cervical lesion; screening; detection; infection; cervix;
XX HPV E4.
XX
OS Human papillomavirus.
XX
XX WO9825145-A1.
XX
XX 11-JUN-1998.
XX
XX 03-DEC-1997; 97WO-GB003321.
XX
XX 03-DEC-1996; 96GB-00025142.
XX
XX 05-SEP-1997; 97GB-00018745.

```

XX PA (MEDI-) MEDICAL RES COUNCIL.
 XX PI Doorbar J;
 XX XX WPI; 1998-333497/29.
 XX PT Detecting papilloma virus infection using molecule binding to E4 protein
 XX PT - useful, e.g. in screening for pre-cancerous cervical lesions and to
 XX PT determine type(s) of human papilloma virus infecting human patients.
 XX PS Claim 11; Page 37; 52pp; English.
 XX XX A new method has been developed for detecting a papilloma virus infection
 CC in an organism. The method comprises: (i) obtaining a sample of cells
 CC from the potential infection site; (ii) contacting the cells with a
 CC molecule binding specifically to papilloma virus E4 protein, and (iii)
 CC monitoring the binding. The method is useful to detect papilloma virus
 CC infections in organisms (especially mammals) and especially HPV
 CC infections (e.g. with HPV16, 18, 33, 35, 45, 51, 56, 58 or 61) in humans.
 CC Papilloma viruses cause epithelial tumours in humans varying in severity
 CC depending on the infection site and HPV type involved. The method is
 CC particularly useful to determine papilloma infection in the mammalian
 CC cervix and especially to screen for pre-cancerous cervical lesions in
 CC humans, since over 90% of cervical carcinoma patients show cervical HPV
 CC infection. It is also useful to determine the type(s) of HPV infection in
 CC a patient, by using a molecule binding specifically to a subset of HPV E4
 CC proteins. This is important, since progression to malignant disease (and
 CC hence clinical prognosis) is dependent on HPV type. Molecules capable of
 CC binding E4 are also useful to target anticancer/antiviral agents capable
 CC of destroying papilloma viruses and/or papilloma virus-infected cells.
 CC The present sequence represents a specifically claimed HPV16 peptide
 CC found in the hydrophilic region
 XX SQ Sequence 12 AA;

Query Match 82.4%; Score 75; DB 2; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KPSPWAPKKHR 15
 |||||
 Db 1 KPSPWAPKKHR 12

RESULT 7
 AAR14288
 ID AAR14288 standard; protein; 14 AA.
 XX AC AAR14288;
 XX DT 02-JAN-1992 (first entry)
 XX DE Seroreactive epitope #2 of HPV16 protein E4.
 XX XX HPV16-dependent human disease; E6; E7; L1.
 XX OS Synthetic.
 XX PN EP451550-A.
 XX PD 16-OCT-1991.
 XX PF 19-MAR-1991; 91EP-00104197.
 XX PR 20-MAR-1990; 90EP-00105222.
 XX PA (BEHW) BEHRINGWERKE AG.
 XX PI Muller M, Gissmann L;
 XX DR WPI; 1991-304643/42.

Query Match 82.4%; Score 75; DB 2; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.0014;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KPSPWAPKKHR 15
 |||||
 Db 1 KPSPWAPKKHR 12

RESULT 7
 AAR14288
 ID AAR14288 standard; protein; 14 AA.
 XX AC AAR14288;
 XX DT 02-JAN-1992 (first entry)
 XX DE Seroreactive epitope #2 of HPV16 protein E4.
 XX XX HPV16-dependent human disease; E6; E7; L1.
 XX OS Synthetic.
 XX PN EP451550-A.
 XX PD 16-OCT-1991.
 XX PF 19-MAR-1991; 91EP-00104197.
 XX PR 20-MAR-1990; 90EP-00105222.
 XX PA (BEHW) BEHRINGWERKE AG.
 XX PI Muller M, Gissmann L;
 XX DR WPI; 1991-304643/42.

PT Sero-active epitope(s) of human papilloma-virus 16 proteins - for use
 PT vaccines and diagnosis.
 XX Claim 1; Page 11; 15pp; English.
 XX CC This is one of two seroreactive epitopes identified from HPV16 protein
 CC E4. Peptides containing this epitope are also claimed and can be used to
 CC generate antibodies to HPV. See AAR14287-R14302 and AAQ14168-Q14171
 XX SQ Sequence 14 AA;

Query Match 74.7%; Score 68; DB 2; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.014;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 KPSPWAPKKHR 15
 |||||
 Db 1 KPSPWAPKKHR 11

RESULT 8
 AAR14287
 ID AAR14287 standard; peptide; 11 AA.
 XX AC AAR14287;
 XX DT 02-JAN-1992 (first entry)
 XX DE Seroreactive epitope #1 of HPV16 protein E4.
 XX HPV16-dependent human disease; E6; E7; L1.
 XX OS Synthetic.
 XX PN EP451550-A.
 XX PD 16-OCT-1991.
 XX PF 19-MAR-1991; 91EP-00104197.
 XX PR 20-MAR-1990; 90EP-00105222.
 XX PA (BEHW) BEHRINGWERKE AG.
 XX PI Muller M, Gissmann L;
 XX DR WPI; 1991-304643/42.
 XX PT Sero-active epitope(s) of human papilloma-virus 16 proteins - for use
 XX PT vaccines and diagnosis.
 XX PS Claim 1; Page 11; 15pp; English.
 XX CC This is one of two seroreactive epitopes identified from HPV16 protein
 XX E4. Peptides containing this epitope are also claimed and can be used to
 XX generate antibodies to HPV. See also AAR14288-R14302 and AAQ14168-Q14171
 XX SQ Sequence 11 AA;

Query Match 72.5%; Score 66; DB 2; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.02;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 IPKSPWAPKK 13
 |||||
 Db 1 IPKSPWAPKK 11

RESULT 9
 AAO03023
 ID AAO03023 standard; protein; 105 AA.
 XX AC AAO03023;

```

XX 06-NOV-2001 (first entry)
DT Human polypeptide SEQ ID NO 16915.
DE
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US-00515126.
XX
XX 18-MAY-2000; 2000US-00577409.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Dmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX N-PSDB; AAI82954.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
XX Claim 20; SEQ ID NO 16915; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 105 AA;
SQ
Query Match 61.5%; Score 56; DB 4; Length 105;
Best Local Similarity 81.8%; Pred. No. 4.1;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 PKPSPWAPKKH 14
DB 82 PCPSPWAPKKH 92

RESULT 10
ADE54638
ID ADE54638 standard; protein; 1053 AA.
XX
XX ADE54638;
AC
XX
XX 29-JAN-2004 (first entry)
DT Human Protein BAA20798, SEQ ID NO 443.
XX
XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX
XX Homo sapiens.
OS

```

```

XX WO2003016475-A2.
XX
XX 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
XX
XX 01-NOV-2001; 2001US-0346382P.
XX
XX 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO ) GEN HOSPITAL CORP.
XX
XX (FARB ) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX
XX WPI; 2003-268312/26.
XX
XX GENBANK; BAA20798.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1053 AA;
SQ
Query Match 59.3%; Score 54; DB 7; Length 1053;
Best Local Similarity 53.8%; Pred. No. 73;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIPKSPWAPKKH 14
DB 890 PLPQSPFMPRRH 902

RESULT 11
ADE54636
ID ADE54636 standard; protein; 1553 AA.
XX
XX ADE54636;
AC
XX
XX 29-JAN-2004 (first entry)
DT Rat Protein AAB66703, SEQ ID NO 441.
XX
XX

```

KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 OS Rattus norvegicus.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

XX 01-NOV-2001; 2001US-0346382P.

XX 26-NOV-2001; 2001US-0333347P.

XX (GCHO) GEN HOSPITAL CORP.

XX (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; AAB66703.

XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which differentially
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1553 AA;

XX Query Match 59.3%; Score 54; DB 7; Length 1553;

XX Best Local Similarity 53.8%; Pred. No. 1.1e+02;

XX Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 PIPKPSWPAPKH 14

Db 890 PLQPSPPEMPREH 902

RESULT 12

AAU48618

ID AAU48618 standard; protein; 135 AA.

XX AC AAU48618;

XX

DT 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #9514.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO2001B1581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIYA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59543.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.

XX Example 1; SEQ ID NO 9813; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 135 AA;

XX Query Match 56.0%; Score 51; DB 4; Length 135;

XX Best Local Similarity 66.7%; Pred. No. 24;

XX Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 RPIPKPSWPAPK 12

Db 69 RPIPMFSSWPFR 80

RESULT 13

ABM45137

ID ABM45137 standard; protein; 135 AA.

XX AC ABM45137;

XX

DT 20-OCT-2003 (first entry)
 DE Propionibacterium acnes predicted ORF-encoded polypeptide #9813.
 XX
 XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine.
 KW
 XX Propionibacterium acnes.
 OS
 XX WO2003033515-A1.
 FN
 XX 24-APR-2003.
 PD
 XX 11-OCT-2002; 2002WO-US032727.
 XX
 XX 15-OCT-2001; 2001US-00978925.
 FR
 XX (CORI-) CORIXA CORP.
 XX
 XX Mitcham JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL;
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 PI Barth B, Vallie-Douglass J;
 XX
 XX WPI; 2003-381789/36.
 DR N-PSDB; ACF64472.
 XX
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 FT or for stimulating an immune response specific for a P. acnes protein.
 PT
 XX
 XX Example 1; SEQ ID NO 9813; 1481pp; English.
 PS
 XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridisation. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open
 CC reading frame) contained within the P. acnes polynucleotides of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 135 AA;
 SQ
 Query Match 56.0%; Score 51; DB 6; Length 135;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 RP1PKPSPWAPK 12
 Db 69 RP1PMPSWEPFR 80
 RESULT 14
 ABB11595

ID XX ABB11595 standard; peptide; 137 AA.
 AC ABB11595;
 XX
 DT 11-JAN-2002 (first entry)
 XX
 DE Human FPM315 homologue, SEQ ID NO:1965.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KW chronic inflammatory condition; proliferative retinopathy;
 KW atherosclerosis; coronary heart disease; arterial ischaemia;
 KW bone disorder; osteoporosis; vascular growth disorder;
 KW tissue regeneration; wound healing; infection; immune disorder;
 KW cell culture; drug screening; gene therapy; antiinflammatory;
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KW cytosstatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KW antifungal; vulnery; antiulcer.
 XX
 OS Homo sapiens.
 XX
 XX WO200157188-A2.
 DN
 XX 09-AUG-2001.
 PD
 XX 05-FEB-2001; 2001WO-US003800.
 XX
 XX 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX WPI; 2001-457740/49.
 DR N-PSDB; ABA08833.
 XX
 XX Human proteins and DNA encoding sequences useful for preventing, treating
 PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
 PT and cancer.
 PT
 XX Claim 20; Page 218; 1963pp; English.
 PS
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with

CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention
 XX
 SQ Sequence 137 AA;

Query Match 55.5%; Score 50.5; DB 4; Length 137;
 Best Local Similarity 64.3%; Pred. No. 29;
 Matches 9; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

Qy 1 RPIPK---PSPWAP 11
 Db 96 RPLPKNARPPSPWVP 109

RESULT 15
 ADE52670
 ID ADE52670 standard; protein; 967 AA.

XX
 AC ADE52670;
 DT 29-JAN-2004 (first entry)
 XX
 DE Human protein SEQ ID 36.
 XX
 KW Human; DNA-binding protein; interferon-activatable protein.
 XX
 OS Homo sapiens.

PN WO2003089466-A1.

XX
 PD 30-OCT-2003.

XX
 PF 18-APR-2003; 2003WO-JP004981.

XX
 PR 19-APR-2002; 2002JP-00117840.

PR 30-APR-2002; 2002JP-00128418.

PR 30-APR-2002; 2002JP-00128779.

PR 04-DEC-2002; 2002JP-00352469.

XX
 PA (RIKE) RIKEN KK.

PA (DNAF-) DNAFORM KK.

PA (MITU) MITSUBISHI CHEM CORP.

XX
 PI Hayashizaki Y, Kamiya M, Kubodera H;

XX
 DR WPI, 2004-011681/01.

XX
 DR N-PSDB; ADE52669.

XX
 PT Proteins with DNA binding activity and substances that affect their

XX
 PT activity or expression, useful for treating associated disorders.

XX
 PS Claim 1; SEQ ID NO 36; 237pp; Japanese.

XX
 CC The present invention relates to novel proteins (ADE52648-ADE52660,

CC ADE52670 and ADE52672) and their coding sequences (ADE52635-ADE52647,

CC ADE52669 and ADE52671). The proteins have a DNA-binding activity or an

XX
 CC interferon-activatable protein (IAP)-like activity.

XX
 SQ Sequence 967 AA;

Query Match 55.5%; Score 50.5; DB 8; Length 967;
 Best Local Similarity 64.3%; Pred. NO. 2e+02;
 Matches 9; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

Qy 1 RPIPK---PSPWAP 11

Db 190 RPLPKNARPPSPWVP 203
 Search completed: May 27, 2004, 16:18:29
 Job time : 43.5385 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:15:59 ; Search time 12.1154 Seconds
(without alignments)
63.918 Million cell updates/sec

Title: US-10-008-524A-167

Perfect score: 91

Sequence: 1 RPIPKPSWPAPKKHR 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfiles!.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	100.0	15	4	US-09-314-268-167
2	91	100.0	25	4	US-09-314-268-4
3	91	100.0	30	2	US-08-934-915-42
4	91	100.0	92	4	US-09-314-268-90
5	79	86.8	20	2	US-08-934-915-40
6	79	86.8	20	2	US-08-934-915-170
7	75	82.4	12	4	US-09-314-268-168
8	68	74.7	35	4	US-09-314-268-117
9	54	59.3	173	4	US-09-252-991A-32711
10	52	57.1	8	4	US-09-314-268-31
11	52	57.1	8	4	US-09-314-268-33
12	52	57.1	8	4	US-09-314-268-37
13	51	56.0	8	4	US-09-314-268-36
14	51	56.0	34	4	US-09-314-268-118
15	50	54.9	8	4	US-09-314-268-34
16	50	54.9	8	4	US-09-314-268-35
17	50	54.9	195	4	US-09-252-991A-22162
18	49	53.8	8	4	US-09-314-268-32
19	49	53.8	325	4	US-09-921-099A-21
20	49	53.8	419	4	US-09-252-991A-22888
21	48	52.7	143	4	US-09-252-991A-21367
22	47	51.6	154	4	US-09-252-991A-25612
23	46	50.5	8	4	US-09-314-268-30
24	46	50.5	81	4	US-09-247-155-125
25	46	50.5	81	4	US-09-663-600A-101
26	46	50.5	81	4	US-09-663-600A-195
27	46	50.5	200	1	US-08-187-829-3

28 46 50.5 200 3 US-09-021-290-3 Sequence 3, Appli
29 46 50.5 200 4 US-09-572-046-3 Sequence 3, Appli
30 46 50.5 201 3 US-08-987-418A-2 Sequence 2, Appli
31 46 50.5 201 3 US-09-343-062-2 Sequence 2, Appli
32 46 50.5 227 6 5498499-2 Patent No. 5498499
33 46 50.5 280 4 US-09-328-352-5299 Sequence 2, Appli
34 46 50.5 993 1 US-08-444-792-2 Sequence 5299, Ap
35 46 50.5 993 1 US-08-445-042-2 Sequence 2, Appli
36 46 50.5 1039 4 US-09-409-648-7 Sequence 7, Appli
37 46 50.5 1039 6 5196511-2 Patent No. 5196511
38 45.5 50.0 254 4 US-09-252-991A-31621 Sequence 31621, A
39 45 49.5 8 4 US-09-314-268-38 Sequence 38, Appli
40 45 49.5 39 4 US-09-314-268-119 Sequence 119, App
41 45 49.5 127 4 US-09-621-976-6400 Sequence 6400, Ap
42 45 49.5 156 4 US-09-252-991A-18484 Sequence 18484, A
43 45 49.5 390 4 US-09-252-991A-17829 Sequence 17829, A
44 45 49.5 493 2 US-08-933-821-2 Sequence 2, Appli
45 45 49.5 493 3 US-08-960-507-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-314-268-167

; Sequence 167, Application US/09314268

; Patent No. 6346377

; GENERAL INFORMATION:

; APPLICANT: Doorbar, John

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

; FILE REFERENCE: 3789/80902

; CURRENT APPLICATION NUMBER: US/09/314,268

; CURRENT FILING DATE: 1999-03-19

; EARLIER APPLICATION NUMBER: 09/314,268

; EARLIER FILING DATE: 1999-05-18

; NUMBER OF SEQ ID NOS: 179

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 167

; LENGTH: 15

; TYPE: PRT

; ORGANISM: Human papillomavirus type 16

US-09-314-268-167

Query Match 100.0%; Score 91; DB 4; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.8e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIPKPSWPAPKKHR 15

Db 1 RPIPKPSWPAPKKHR 15

RESULT 2

US-09-314-268-4

; Sequence 4, Application US/09314268

; Patent No. 6346377

; GENERAL INFORMATION:

; APPLICANT: Doorbar, John

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

; FILE REFERENCE: 3789/80902

; CURRENT APPLICATION NUMBER: US/09/314,268

; CURRENT FILING DATE: 1999-03-19

; EARLIER APPLICATION NUMBER: 09/314,268

; EARLIER FILING DATE: 1999-05-18

; NUMBER OF SEQ ID NOS: 179

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 25

; TYPE: PRT

; ORGANISM: Human papillomavirus type 16

US-09-314-268-4

```
Query Match      100.0%; Score 91; DB 4; Length 25;
Best Local Similarity 100.0%; Pred. No. 3.1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKHR 15
   |||||
Db 1 RPIPKPSWAPKKHR 15

RESULT 3
US-08-934-915-42
; Sequence 42, Application US/08934915
; Patent No. 5932412
; GENERAL INFORMATION:
; APPLICANT: DILLNER, JOAKIM
; APPLICANT: DILLNER, LENA
; TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
; TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,
; TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
; TITLE OF INVENTION: USEFUL IN IMMUNOASSAY FOR
; TITLE OF INVENTION: DIAGNOSTIC PURPOSES
; NUMBER OF SEQUENCES: 193
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MASON & ASSOCIATES, P.A.
; STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
; CITY: CLEARWATER
; STATE: FLORIDA
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 3.0
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,915
; FILING DATE: 22-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/949,836
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: LOUISE A. Foutch
; REGISTRATION NUMBER: 37,133
; REFERENCE/DOCKET NUMBER: 1946.6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 813-538-3800
; TELEFAX: 813-538-3820
; TELEX:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-934-915-42

Query Match      100.0%; Score 91; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.7e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKHR 15
   |||||
Db 7 RPIPKPSWAPKKHR 21

RESULT 4
US-09-314-268-90
; Sequence 90, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
```

```
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-09-314-268-90

Query Match      100.0%; Score 91; DB 4; Length 92;
Best Local Similarity 100.0%; Pred. No. 1.1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKHR 15
   |||||
Db 26 RPIPKPSWAPKKHR 40

RESULT 5
US-08-934-915-40
; Sequence 40, Application US/08934915
; Patent No. 5932412
; GENERAL INFORMATION:
; APPLICANT: DILLNER, JOAKIM
; APPLICANT: DILLNER, LENA
; APPLICANT: CHENG, HWEI-MING
; TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
; TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,
; TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
; TITLE OF INVENTION: USEFUL IN IMMUNOASSAY FOR
; TITLE OF INVENTION: DIAGNOSTIC PURPOSES
; NUMBER OF SEQUENCES: 193
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MASON & ASSOCIATES, P.A.
; STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
; CITY: CLEARWATER
; STATE: FLORIDA
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 3.0
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,915
; FILING DATE: 22-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/949,836
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: LOUISE A. Foutch
; REGISTRATION NUMBER: 37,133
; REFERENCE/DOCKET NUMBER: 1946.6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 813-538-3800
; TELEFAX: 813-538-3820
; TELEX:
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-934-915-40

Query Match      86.8%; Score 79; DB 2; Length 20;
```


Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IPKSPWAPKKHR 15
| | | | | | | | | | | | | | |
Db 1 IPKSPWAPKKHR 13

RESULT 6

US-08-934-915-170
; Sequence 170, Application US/08934915
; Patent No. 5932412
; GENERAL INFORMATION:
; APPLICANT: DILLNER, JOAKIM
; APPLICANT: DILLNER, LENA
; APPLICANT: CHENG, HWEI-MING
; TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
; TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,
; TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
; TITLE OF INVENTION: USEFUL IN IMMUNOASSAY FOR
; TITLE OF INVENTION: DIAGNOSTIC PURPOSES
; NUMBER OF SEQUENCES: 193
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MASON & ASSOCIATES, P.A.
; STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
; CITY: CLEARWATER
; STATE: FLORIDA
; COUNTRY: U.S.A.

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 3.0
; SOFTWARE: Microsoft word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,915
; FILING DATE: 22-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/949,836
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: LOUISE A. Foutch
; REGISTRATION NUMBER: 37,133
; REFERENCE/DOCKET NUMBER: 1946.6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 813-538-3800
; TELEFAX: 813-538-3820
; TELEX:

; INFORMATION FOR SEQ ID NO: 170:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-934-915-170

Query Match 86.8%; Score 79; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IPKSPWAPKKHR 15
| | | | | | | | | | | | | | |
Db 1 IPKSPWAPKKHR 13

RESULT 7

US-09-314-268-168
; Sequence 168, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES

; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 168
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-09-314-268-168

Query Match 82.4%; Score 75; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PKSPWAPKKHR 15
| | | | | | | | | | | | | | |
Db 1 PKSPWAPKKHR 12

RESULT 8

US-09-314-268-117
; Sequence 117, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 117
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-09-314-268-117

Query Match 74.7%; Score 68; DB 4; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 KPSPWAPKKHR 15
| | | | | | | | | | | | | | |
Db 1 KPSPWAPKKHR 11

RESULT 9

US-09-252-991A-32711
; Sequence 32711, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32711
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32711

Query Match 59.3%; Score 54; DB 4; Length 173;
Best Local Similarity 72.7%; Pred. No. 2.3;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPIKPSPWP 11
|||:||||
Db 158 RVPRIISPWP 168

RESULT 10
US-09-314-268-31
; Sequence 31, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-09-314-268-31

Query Match 57.1%; Score 52; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PIRKPSPW 9
|||||
Db 1 PIRKPSPW 8

RESULT 11
US-09-314-268-33
; Sequence 33, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-09-314-268-33

Query Match 57.1%; Score 52; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PKPSPWP 11
|||||

Db 1 PKPSPWP 8

RESULT 12
US-09-314-268-37
; Sequence 37, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-09-314-268-37

Query Match 57.1%; Score 52; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 PWAPKXHR 15
|||||
Db 1 PWAPKXHR 8

RESULT 13
US-09-314-268-36
; Sequence 36, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-09-314-268-36

Query Match 56.0%; Score 51; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 SPWAPKXH 14
|||||
Db 1 SPWAPKXH 8

RESULT 14
US-09-314-268-118
; Sequence 118, Application US/09314268
; Patent No. 6346377

```
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 118
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Human papillomavirus type 35
US-09-314-268-118

Query Match      56.0%; Score 51; DB 4; Length 34;
Best Local Similarity 72.7%; Pred. No. 1.1;
Matches      8; Conservative      2; Mismatches      1; Indels      0; Gaps      0;

QY      5 KPSPWAPKHR 15
Db      1 KPAPWAPQKPR 11

RESULT 15
US-09-314-268-34
; Sequence 34, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-314-268-34

Query Match      54.9%; Score 50; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY      5 KPSPWAPK 12
Db      1 KPSPWAPK 8
```

Search completed: May 27, 2004, 16:22:37
Job time : 13.1154 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:21:04 ; Search time 31.7308 Seconds
(without alignments)
132.009 Million cell updates/sec

Title: US-10-008-524A-167
Perfect score: 91
Sequence: 1 RPIPKSPWAPKKHR 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1151071 seqs, 279249464 residues

Total number of hits satisfying chosen parameters: 1151071

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	100.0	15	14	US-10-008-524A-167
2	91	100.0	25	14	US-10-008-524A-4
3	91	100.0	26	15	US-10-350-719-4
4	91	100.0	92	14	US-10-008-524A-90
5	91	100.0	92	15	US-10-350-719-90
6	82	90.1	15	15	US-10-350-719-167
7	75	82.4	12	14	US-10-008-524A-168
8	75	82.4	12	15	US-10-350-719-168
9	68	74.7	35	14	US-10-008-524A-117
10	68	74.7	35	15	US-10-350-719-117
11	55	60.4	113	12	US-10-424-599-254084
12	53	58.2	55	14	US-10-029-386-31681
13	52	57.1	8	14	US-10-008-524A-31
14	52	57.1	8	14	US-10-008-524A-33
15	52	57.1	8	14	US-10-008-524A-37

16	52	57.1	8	15	US-10-350-719-31	Sequence 31, Appl
17	52	57.1	8	15	US-10-350-719-33	Sequence 33, Appl
18	52	57.1	8	15	US-10-350-719-37	Sequence 37, Appl
19	51	56.0	8	14	US-10-008-524A-36	Sequence 36, Appl
20	51	56.0	8	15	US-10-350-719-36	Sequence 36, Appl
21	51	56.0	34	14	US-10-008-524A-118	Sequence 118, App
22	51	56.0	34	15	US-10-350-719-118	Sequence 118, App
23	50.5	55.5	137	12	US-10-276-774-1965	Sequence 1965, Ap
24	50	54.9	8	14	US-10-008-524A-34	Sequence 34, Appl
25	50	54.9	8	14	US-10-008-524A-35	Sequence 35, Appl
26	50	54.9	8	15	US-10-350-719-34	Sequence 34, Appl
27	50	54.9	8	15	US-10-350-719-35	Sequence 35, Appl
28	49	53.8	8	14	US-10-008-524A-32	Sequence 32, Appl
29	49	53.8	8	15	US-10-350-719-32	Sequence 32, Appl
30	49	53.8	93	12	US-10-424-599-219309	Sequence 219309,
31	49	53.8	142	15	US-10-131-487A-125	Sequence 125, App
32	49	53.8	168	15	US-10-131-487A-192	Sequence 192, App
33	49	53.8	234	12	US-10-424-599-266133	Sequence 266133,
34	48	52.7	783	15	US-10-440-559-6	Sequence 6, Appli
35	48	52.7	783	15	US-10-440-559-6	Sequence 6, Appli
36	48	52.7	1590	14	US-10-180-326-1	Sequence 1, Appli
37	48	52.7	8	14	US-10-008-524A-30	Sequence 30, Appl
38	46	50.5	8	15	US-10-350-719-30	Sequence 30, Appl
39	46	50.5	81	10	US-09-903-190-125	Sequence 125, App
40	46	50.5	81	14	US-10-319-763-101	Sequence 101, App
41	46	50.5	81	14	US-10-319-763-195	Sequence 195, App
42	46	50.5	81	14	US-10-424-599-264398	Sequence 264398,
43	46	50.5	114	14	US-10-029-386-30822	Sequence 30822,
44	46	50.5	200	9	US-09-972-137-3	Sequence 3, Appli
45	46	50.5				

ALIGNMENTS

RESULT 1

US-10-008-524A-167
; Sequence 167, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 167
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-167

Query Match 100.0%; Score 91; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00014;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIPKSPWAPKKHR 15
| | | | | | | | | | | | | | |
Db 1 RPIPKSPWAPKKHR 15

RESULT 2

US-10-008-524A-4
; Sequence 4, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES


```
RESULT 7
US-10-008-524A-168
; Sequence 168, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 168
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-168

Query Match      82.4%; Score 75; DB 14; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 KPSPWAPKKHR 15
Db      1 KPSPWAPKKHR 12
|||||

RESULT 8
US-10-350-719-168
; Sequence 168, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 168
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-350-719-168

Query Match      82.4%; Score 75; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 KPSPWAPKKHR 15
Db      1 KPSPWAPKKHR 12
|||||

RESULT 9
US-10-008-524A-117
; Sequence 117, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
```

```
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 117
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-117

Query Match      74.7%; Score 68; DB 14; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 KPSPWAPKKHR 15
Db      1 KPSPWAPKKHR 11
|||||

RESULT 10
US-10-350-719-117
; Sequence 117, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 117
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-350-719-117

Query Match      74.7%; Score 68; DB 15; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 KPSPWAPKKHR 15
Db      1 KPSPWAPKKHR 11
|||||

RESULT 11
US-10-424-599-254084
; Sequence 254084, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 254084
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
```

; OTHER INFORMATION: Clone ID: PAT_MRT3847_71461C.1.pcp
US-10-424-599-254084

Query Match 60.4%; Score 55; DB 12; Length 113;
Best Local Similarity 76.9%; Pred. No. 17;
Matches 10; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RPIPKPSWAPKK 13
||| |||||
Db 34 RPFSPSPWAPKK 46

RESULT 12

US-10-029-386-31681
; Sequence 31681, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31681
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008974.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.55
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8
; OTHER INFORMATION: SWISSPROT HIT: Q28084, EVALUATE 1.20e+00
US-10-029-386-31681

Query Match 58.2%; Score 53; DB 14; Length 55;
Best Local Similarity 57.1%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 RPIPKPSWAPKK 14
||| |||||
Db 7 RVPGASWPFGSH 20

RESULT 13

US-10-008-524A-31
; Sequence 31, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR FILING DATE: 09/314,268
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 31
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-10-008-524A-31

Query Match 57.1%; Score 52; DB 14; Length 8;

Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PIPKPSW 9
|||||
Db 1 PIPKPSW 8

RESULT 14

US-10-008-524A-33
; Sequence 33, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR FILING DATE: 09/314,268
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-10-008-524A-33

Query Match 57.1%; Score 52; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PKPSPWAP 11
|||||
Db 1 PKPSPWAP 8

RESULT 15

US-10-008-524A-37
; Sequence 37, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR FILING DATE: 09/314,268
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-10-008-524A-37

Query Match 57.1%; Score 52; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 PWAPKKHR 15
|||||
Db 1 PWAPKKHR 8

Search completed: May 27, 2004, 16:34:18
Job time : 31.7308 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:13:48 ; Search time 10.0962 Seconds
(without alignments)
142.913 Million cell updates/sec

Title: US-10-008-524A-167

Sequence: 1 RPIPKPSFWAPKKHR 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	100.0	95	1 W4WLHS	E4 protein - human
2	74	81.3	96	1 W4WL35	E4 protein - human
3	61	67.0	87	1 W4WL51	E4 protein - human
4	56	61.5	102	1 W4WL31	E4 protein - human
5	54	59.3	330	2 T05717	probable extensin
6	54	59.3	1553	2 T03301	rab3 effector prot
7	51	56.0	1060	2 A10201	beta-galactosidase
8	49	53.8	325	2 S57977	COOH zinc finger p
9	48	52.7	88	1 W4WL18	E4 protein - human
10	48	52.7	721	2 C94677	probable membrane
11	48	52.7	992	2 T38817	hypothetical prote
12	47	51.6	94	1 W4WL39	E4 protein - human
13	47	51.6	378	1 A40004	histidine decarbox
14	47	51.6	846	2 S52418	GTP-binding regula
15	47	51.6	925	2 T07713	probable ABC-type
16	46	50.5	230	2 S72714	Lepb1170 F2.64 pro
17	46	50.5	474	2 T34193	G protein-coupled
18	46	50.5	688	2 B64103	glycine-tRNA ligas
19	46	50.5	1007	2 E72489	hypothetical prote
20	46	50.5	1039	2 A34269	integrin alpha-2b
21	45	49.5	283	2 S13383	hydroxyproline-ric
22	45	49.5	389	2 T05782	hypothetical prote
23	45	49.5	406	2 T20330	nonhistone chromos
24	44	48.4	105	1 NSCH44	pollen allergen pI
25	44	48.4	308	2 A38582	hypothetical prote
26	44	48.4	372	2 T01600	hypothetical prote
27	44	48.4	798	2 T50514	hypothetical prote
28	44	48.4	981	2 T16060	hypothetical prote
29	43.5	47.8	143	2 D93148	hypothetical prote

ALIGNMENTS

RESULT 1

W4WLHS

E4 protein - human papillomavirus type 16

C:Species: human papillomavirus type 16

C>Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 20-Aug-1999

C:Accession: A22355; T10425

R:Seedorf, K.; Krammer, G.; Durst, M.; Suhai, S.; Rowekamp, W.G.

Virology 145, 181-185, 1985

A>Title: Human papillomavirus type 16 DNA sequence.

A:Reference number: A22355; MUID:85246220; PMID:2990099

A:Accession: A22355

A:Molecule type: DNA

A:Residues: 1-95 <SEE>

A:Cross-references: GB:K02718; NID:G333031; PIDN:AAA46937.1; PID:G459913

J.Kennedy, J.M.; Haddow, J.K.; Clements, J.B.

R.Virol. 45, 2093-2097, 1991

A>Title: A negative element in the human papillomavirus type 16 genome acts at the lev

A:Reference number: Z17014; MUID:91162763; PMID:1848319

A:Accession: T10425

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-95 <KEN>

A:Cross-references: EMBL:K02718; NID:G333031; PIDN:AAA46937.1; PID:G459913

C:Genetics:

A:Gene: E4

C:Superfamily: papillomavirus E4 protein

C:Keywords: early protein

Query Match 100.0%; Score 91; DB 1; Length 95;

Best Local Similarity 100.0%; Pred.No. 1.3e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIPKPSFWAPKKHR 15

Db 29 RPIPKPSFWAPKKHR 43

RESULT 2

W4WL35

E4 protein - human papillomavirus type 35

C:Species: human papillomavirus type 35

A>Note: host Homo sapiens (man)

C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 27-Jan-1995

C:Accession: C40824

R:Marich, J.E.; Pontsler, A.V.; Rice, S.M.; McGraw, K.A.; Dubensky, T.W.

Virology 186, 770-776, 1992

A>Title: The phylogenetic relationship and complete nucleotide sequence of human papill

A:Reference number: A40824; MUID:92124753; PMID:1310198

A:Accession: C40824

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-96 <MAP>

A;Cross-references: GB:M74117
C;Superfamily: papillomavirus E4 protein
C;Keywords: early protein

Query Match 81.3%; Score 74; DB 1; Length 96;
Best Local Similarity 80.0%; Pred. No. 0.0026;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKR 15
|||||:|||||
DB 29 RPIKPAPWAPQKPR 43

RESULT 3

W4WL51

E4 protein - human papillomavirus type 51
C;Species: human papillomavirus type 51
A;Note: host Homo sapiens (man)

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 27-Jan-1995
C;Accession: C40415

R;Lungu O.; Crum, C.P.; Silverstein, S.J.

J. Virol. 65, 4216-4225, 1991

A;Title: Biologic properties and nucleotide sequence analysis of human papillomavirus type 51
A;Reference number: A40415; PMID:91303675; PMID:1649326

A;Accession: C40415

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-87 <LUN>

A;Cross-references: GB:M62877

C;Superfamily: papillomavirus E4 protein

C;Keywords: early protein

Query Match 67.0%; Score 61; DB 1; Length 87;
Best Local Similarity 73.3%; Pred. No. 0.13;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKR 15
|||||:|||||
DB 24 RPIPLPAPWAPKKPR 38

RESULT 4

W4WL31

E4 protein - human papillomavirus type 31

C;Species: human papillomavirus type 31

A;Note: host Homo sapiens (man)

C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999

C;Accession: E32444

R;Goldsborough, M.D.; Disilvestre, D.; Temple, G.F.; Lorincz, A.T.

Virolgy 171, 306-311, 1989

A;Title: Nucleotide sequence of human papillomavirus type 31: a cervical neoplasia-associated

A;Reference number: A94398; PMID:89299478; PMID:2545036

A;Accession: E32444

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-102 <GOL>

A;Cross-references: GB:J04353; NID:9333048; PIDN:AAA46949.1; PID:9459915

A;Note: in GenBank entry PPH31A the initiation codon UUG for residue 1 is translated as

C;Superfamily: papillomavirus E4 protein

C;Keywords: early protein

Query Match 61.5%; Score 56; DB 1; Length 102;
Best Local Similarity 81.8%; Pred. No. 0.71;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 IPKPSWAPKKR 13
|||||:|||||
DB 32 IPKPAPWAPVK 42

RESULT 5

T05717

probable extensin - barley (fragment)

C;Species: Hordeum vulgare (barley)
C;Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 20-Jun-2000
C;Accession: T05717

R;Doan, D.N.P.; Sturaro, M.; Olsen, O.A.

submitted to the EMBL Data Library, July 1997

A;Description: Characterization of a nuclear cDNA encoding a probable extensin from de

A;Reference number: Z15429

A;Accession: T05717

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 1-330 <DOA>

A;Cross-references: EMBL:Z98204; PIDN:CAB10894.1

C;Genetics:

A;Gene: ex1

A;Map position: 2

C;Superfamily: hydroxyproline-rich glycoprotein

Query Match 59.3%; Score 54; DB 2; Length 330;
Best Local Similarity 75.0%; Pred. No. 4.1;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPK 12
:|:|:|:|:|
DB 200 KEVPKPSPPAPK 211

RESULT 6

T03301

rab3 effector protein Rim - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999

C;Accession: T03301

R;Wang, Y.; Okamoto, M.; Schmitz, F.; Hofmann, K.; Sudhof, T.C.

Nature 388, 593-598, 1997

A;Title: Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion.

A;Reference number: Z14897; PMID:97394473; PMID:9252191

A;Accession: T03301

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: mRNA

A;Residues: 1-1553 <WAN>

A;Cross-references: EMBL:AF007836; NID:G2317777; PIDN:AAB66703.1; PID:G2317778

A;Experimental source: tissue-type brain

C;Genetics:

A;Note: RIM

C;Function:

A;Description: Rim protein is proposed as Rab3-dependent regulator of synaptic-vesicle

C;Keywords: GTP binding; zinc finger

Query Match 59.3%; Score 54; DB 2; Length 1553;
Best Local Similarity 53.8%; Pred. No. 18;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 RPIPKPSWAPKKH 14
|:|:|:|:|:|
DB 890 FLPOFSPFMPRRH 902

RESULT 7

AI0201

Beta-galactosidase (EC 3.2.1.23) [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001

C;Accession: AI0201

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; PMID:21470413; PMID:11586360

A;Accession: AI0201

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1060 <KUR>

A;Cross-references: GB:AL590842; PIDN:CA090476.1; PID:gl5979691; GSPDB:GN00175
C;Genetics:
A;Gene: lacZ
C;Superfamily: beta-galactosidase
C;Keywords: glycosidase; hydrolase

Query Match 56.0%; Score 51; DB 2; Length 1060;
Best Local Similarity 58.3%; Pred. No. 31;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 PKPSPWAPKKR 15
||:|||||
Db 723 PKETWSPAQR 734

RESULT 8

CCCH zinc finger protein CTH1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YD8358.07c; protein YDR151C
C;Species: Saccharomyces cerevisiae
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
C;Accession: S57977; JCS001
R;Murphy, L.; Richards, C.; Harris, D.
submitted to the EMBL Data Library, July 1995
A;Reference number: S57971
A;Accession: S57977
A;Molecule type: DNA
A;Residues: 1-325 <MUR>
A;Cross-references: EMBL:Z50046; NID:G899393; PIDN:CAA90373.1; PID:G899400; MIPS:YDR151C
A;Experimental source: strain AB972
R;Thompson, M.J.; Lai, W.S.; Taylor, G.A.; Blackshear, P.J.
Gene 174, 225-233, 1996
A;Title: Cloning and characterization of two yeast genes encoding members of the CCCH cl
A;Reference number: JCS001; MUID:97045817; PMID:8890739
A;Accession: JCS001
A;Molecule type: DNA
A;Residues: 1-141, 'RV', 144-325 <THO>
A;Cross-references: GB:I42133; NID:gl020082; PIDN:BA039897.1; PID:gl020083
C;Comment: This protein belongs to the CCCH-type zinc finger protein family, and is a su
C;Genetics:
A;Gene: SGD:CTH1
A;Cross-references: SGD:S0002558; MIPS:YDR151c
A;Map position: 4R
C;Keywords: zinc finger
F;132-136/Region: nuclear location signal

Query Match 53.8%; Score 49; DB 2; Length 325;
Best Local Similarity 70.0%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 PKPSPWAPKK 13
|:|||||
Db 79 PRSPWLPK 88

RESULT 9

W4WL18
E4 protein - human papillomavirus type 18
C;Species: human papillomavirus type 18
C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
C;Accession: E26251
R;Cole, S.T.; Danos, O.
J. Mol. Biol. 193, 599-608, 1987
A;Title: Nucleotide sequence and comparative analysis of the human papillomavirus type 1
A;Reference number: A92937; MUID:87283882; PMID:3039146
A;Accession: E26251
A;Molecule type: DNA
A;Residues: 1-88 <COL>
A;Cross-references: GB:X05015; NID:G60975; PIDN:CAA28668.1; PID:G60980
C;Superfamily: papillomavirus E4 protein
C;Keywords: early protein

Query Match 52.7%; Score 48; DB 1; Length 88;

Best Local Similarity 63.6%; Pred. No. 7.4;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 IPKPSWPAPKK 13
||:|||||
Db 28 IPAPCPWAPQR 38

RESULT 10

C84677
probable membrane transporter [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84677
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84677
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-721 <STO>
A;Cross-references: GB:AE002093; NID:G3860251; PIDN:AAC73019.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g27810
A;Map position: 2

Query Match 52.7%; Score 48; DB 2; Length 721;
Best Local Similarity 63.6%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RPpPKPSPWAP 11
:|:|||||
Db 7 KPGPKPGFWPP 17

RESULT 11

T38817
hypothetical protein SPAC4F10.13c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38817
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21813
A;Accession: T38817
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-992 <CON>
A;Cross-references: EMBL:Z98980; NID:el060691; PIDN:CAB11716.1; GSPDB:GN00066; SPDB:SPA
A;Experimental source: strain 972h-; cosmid c4f10
C;Genetics:
A;Gene: SPDB:SPAC4F10.13c
A;Map position: 1
A;Introns: 13/2

Query Match 52.7%; Score 48; DB 2; Length 992;
Best Local Similarity 69.2%; Pred. No. 74;
Matches 9; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 4 PKPSPW--APKKH 14
|||||
Db 622 PKPSPWKSUPPKH 634

RESULT 12

W4WL39
E4 protein - human papillomavirus type 39
C;Species: human papillomavirus type 39
A;Note: host Homo sapiens (man)
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 27-Jan-1995

C;Accession: E38502
R;Volpers, C.; Streack, R.E.
Viology 181, 419-423, 1991
A;Title: Genome organization and nucleotide sequence of human papillomavirus type 39.
A;Reference number: A38502; MUID:91135017; PMID:1847266
A;Accession: E38502
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-94 <VOL>
A;Cross-references: EMBL:M38185
C;Superfamily: papillomavirus E4 protein
C;Keywords: early protein

Query Match 51.6%; Score 47; DB 1; Length 94;
Best Local Similarity 69.2%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RPIPKSPWAPKK 13
|||||
Db 29 RPIPPQPHAPKK 41
|||||

RESULT 13

A40004
histidine decarboxylase (EC 4.1.1.22) - Enterobacter aerogenes
C;Species: Enterobacter aerogenes
C;Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 18-Jun-1999
C;Accession: A40004
R;Kamath, A.V.; Vaaler, G.L.; Snell, E.E.
J. Biol. Chem. 266, 9432-9437, 1991
A;Title: Pyridoxal phosphate-dependent histidine decarboxylases. Cloning, sequencing, and enzymes.

A;Reference number: A40004; MUID:91236707; PMID:2033044

A;Accession: A40004

A;Status: not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-378 <KAM>

A;Cross-references: GB:M62745; NID:9435593; PIDN:AAA24802.1; PID:g435594

C;Superfamily: Klebsiella histidine decarboxylase

C;Keywords: carbon-carbon lyase; carboxy-lyase; phosphoprotein; pyridoxal phosphate

F;233/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 51.6%; Score 47; DB 1; Length 378;
Best Local Similarity 72.7%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 KPSPWAPKKH 14
|||||
Db 329 KPSEWVWKX 339
|||||

RESULT 14

SS2418

GTP-binding regulatory protein Gs alpha-XL chain - rat

N;Alternate names: G protein XL-alpha-s

C;Species: Rattus norvegicus (Norway rat)

C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 02-Feb-2001

C;Accession: SS2418

R;Kehlenbach, R.H.; Matthey, J.; Huttner, W.B.

Nature 372, 804-809, 1994

A;Title: XL-alpha-s is a new type of G protein.

A;Reference number: SS2418; MUID:95089824; PMID:7997272

A;Accession: SS2418

A;Molecule type: mRNA

A;Residues: 1-846 <KEH>

A;Cross-references: EMBL:X84047; NID:9642267; PIDN:CAA58866.1; PID:9642268

R;Kehlenbach, R.H.; Matthey, J.; Huttner, W.B.

Nature 375, 253, 1995

A;Title: Correction: Xlalphas is a new type of G protein.

A;Reference number: SS891

A;Contents: annotation; assignment of start codon

A;Note: experimental data from this paper suggest that the translation is initiated at P

C;Keywords: GTP binding; nucleotide binding; P-loop; signal transduction

F;132-846/Product: GTP-binding regulatory protein Gs alpha-XL chain #status experimental
F;499-506/Region: nucleotide-binding motif A (P-loop)

F;744-747/Region: GTP-binding NKXD motif

Query Match 51.6%; Score 47; DB 2; Length 846;
Best Local Similarity 53.3%; Pred. No. 87;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 RPIPKSPWAPKKHR 15
|||||
Db 350 RPAPPSAWPDXYR 364
|||||

RESULT 15

T07713

probable ABC-type transport protein T23J7.70 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 22-Oct-1999

C;Accession: T07713

R;Quetier, F.; Choisme, N.; Robert, C.; Brottier, P.; Cattolico, L.; Artig

submitted to the Protein Sequence Database, April 1999

A;Reference number: Z15793

A;Accession: T07713

A;Molecule type: DNA

A;Residues: 1-925 <QUE>

A;Cross-references: EMBL:AL049746; GSPDB:GN00061; ATSP:T23J7.70

A;Experimental source: cultivar Columbia; BAC clone T23J7

C;Genetics:

A;Gene: ATSP:T23J7.70

A;Map position: 3

A;Introns: 135/1; 161/3; 210/1; 256/1; 297/3; 350/3; 405/1; 447/1; 460/1; 586/3; 653/3;

C;Keywords: ATP; P-loop

Query Match 51.6%; Score 47; DB 2; Length 925;
Best Local Similarity 77.8%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 IPKSPWAP 11
|||||
Db 84 IPKQPWPP 92
|||||

Search completed: May 27, 2004, 16:21:44

Job time : 11.0962 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:03:58 ; Search time 6.92308 Seconds
(without alignments)
112.819 Million cell updates/sec

Title: US-10-008-524A-167
Perfect score: 91
Sequence: 1 RPIPKSPWAPKKHR 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	100.0	95	VE4_HPV16	P06922 human papil
2	74	81.3	96	VE4_HPV35	P27224 human papil
3	61	67.0	87	VE4_HPV51	P26548 human papil
4	56	61.5	102	VE4_HPV31	P27384 human papil
5	54	59.3	1615	RIM1_RAT	Q9jir4 rattus norv
6	54	59.3	1632	RIM1_HUMAN	Q86ur5 homo sapien
7	49	53.8	325	CTH1_YEAST	P47976 saccharomyc
8	48.5	53.3	1513	GRLF_HUMAN	Q9nrv4 homo sapien
9	48	52.7	88	VE4_HPV18	P06731 human papil
10	48	52.7	1530	RIM2_MOUSE	Q9eqz7 mus musculu
11	47	51.6	94	VE4_HPV39	P24831 human papil
12	47	51.6	377	DCHS_ENTAE	P28577 enterobacte
13	46	50.5	454	CT81_HUMAN	Q9h1q7 homo sapien
14	46	50.5	688	SYGB_HAFIN	P43822 haemophilus
15	46	50.5	689	SYGB_PASMU	P57905 pasteurella
16	46	50.5	755	MTS1_HUMAN	Q43312 homo sapien
17	46	50.5	759	MTS1_MOUSE	Q9rlr4 mus musculu
18	46	50.5	1039	ITAB_HUMAN	P08514 mus musculu
19	45	49.5	153	PE1_ANOGA	Q76217 anopheles g
20	45	49.5	219	MGH1_HUMAN	Q9h213 homo sapien
21	45	49.5	283	EXTN_SORBI	P24152 sorghum bic
22	45	49.5	361	IHA_TRIVU	Q77755 trichosurus
23	45	49.5	461	Y514_HUMAN	Q60269 homo sapien
24	45	49.5	493	ANL2_HUMAN	Q9ukn9 homo sapien
25	45	49.5	1188	RIM2_HUMAN	Q9uq26 homo sapien
26	45	49.5	1555	RIM2_RAT	Q9j1s1 rattus norv
27	44	48.4	104	HG15_CHICK	P12902 gallus gall
28	44	48.4	174	P122_SECC	Q06036 secale cere
29	44	48.4	308	MP5A_LOLFR	Q40240 lolium pere
30	44	48.4	670	ZN16_HUMAN	P17020 homo sapien
31	43.5	47.8	814	TOP1_SCHPO	P07799 schizosacch
32	43	47.3	111	UL91_HCMVA	P16797 human cytom
33	43	47.3	128	VDBP_CERV	P05398 carnation e

34	43	47.3	166	1	RRS1_SCHPO	O59678 schizosacch
35	43	47.3	205	1	YNP2_CABEL	P34555 caenorhabdi
36	43	47.3	251	1	Y786_BRAJA	Q89wa6 bradyrhizob
37	43	47.3	305	1	AMP3_SYNV3	P53581 synechocyst
38	43	47.3	393	1	TRMB_HELPY	Q25443 helicobacte
39	43	47.3	400	1	TRMB_HELPJ	Q32196 helicobacte
40	43	47.3	438	1	NQO1_THETH	Q56222 thermus the
41	43	47.3	438	1	TRF1_CRIGR	O55036 cricetulus
42	43	47.3	439	1	TRF1_HUMAN	P54274 homo sapien
43	43	47.3	459	1	MIG6_RAT	O50432 rattus norv
44	43	47.3	461	1	MIG6_MOUSE	Q99jz7 mus musculu
45	43	47.3	478	1	BM3B_HUMAN	P55107 homo sapien

ALIGNMENTS

RESULT 1
VE4_HPV16
ID VE4_HPV16 STANDARD; PRT; 95 AA.
AC P06922;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85246220; PubMed=2990099;
RA Seedorf K., Krammer G., Durst M., Suhai S., Rowekamp W.G.;
RT "Human papillomavirus type 16 DNA sequence.";
RL Virology 145:181-185(1985).
RN [2]
RP SEQUENCE OF 6-95 FROM N.A.
RX MEDLINE=90218027; PubMed=2157796;
RA Schneider-Maunoury S., Pehau-Arnaudet G., Breitburd F., Orth G.;
RT "Expression of the human papillomavirus type 16 genome in SK-v cells,
a line derived from a vulvar intraepithelial neoplasia.";
RL J. Gen. Virol. 71:809-817(1990).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02718; AAA46937.1; -;
CC EMBL; D00735; BAA00634.1; -;
CC PIR; A23255; W4WLHS.
CC InterPro; IPR003861; Papilloma_E4.
CC Pfam; PF02711; Pap_E4; 1.
CC Early protein.
SQ SEQUENCE 95 AA; 10594 MW; AED4269D177307CE CRC64;

Query Match 100.0%; Score 91; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPIPKSPWAPKKHR 15

Db 29 RPIPKSPWAPKKHR 43

RESULT 2

VE4_HPV35

ID VE4_HPV35 STANDARD; PRT; 96 AA.

AC P27224;

DT 01-AUG-1992 (Rel. 23, Created)

```
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 35.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10587;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92124753; PubMed=1310198;
RA Marich J.E., Pontsler A.V., Rice S.M., McGraw K.A., Dubensky T.W.;
RT "The phylogenetic relationship and complete nucleotide sequence of
human papillomavirus type 35."
RL Virology 186:770-776(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M74117; AAA46965.2; -
DR PIR; C40824; W4WL35.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.
SQ SEQUENCE 96 AA; 10597 MW; AB4524418CD26F7C CRC64;

Query Match 81.3%; Score 74; DB 1; Length 96;
Best Local Similarity 80.0%; Pred. No. 0.0021;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKHR 15
  |||||:||||:|
Db 29 RPIKPAPWAPQKPR 43

RESULT 3
VE4_HPV51
ID -VE4 HPV51 STANDARD; PRT; 87 AA.
AC P26548;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 51.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10595;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91303675; PubMed=1649326;
RA Lungu O., Crum C.P., Silverstein S.J.;
RT "Biologic properties and nucleotide sequence analysis of human
papillomavirus type 51."
RL J. Virol. 65:4216-4225(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M62877; -; NOT ANNOTATED CDS.
DR PIR; C40415; W4WL51.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.
SQ SEQUENCE 87 AA; 9941 MW; 5F3DC38F86BF3990 CRC64;

Query Match 67.0%; Score 61; DB 1; Length 87;
Best Local Similarity 73.3%; Pred. No. 0.0094;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPIPKPSWAPKKHR 15
  |||||:|||||
Db 24 RPIPLPAPWAPKKPR 38

RESULT 4
VE4_HPV31
ID -VE4 HPV31 STANDARD; PRT; 102 AA.
AC P17384;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 31.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10585;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89299478; PubMed=2545036;
RA Goldsborough M.D., Disilvestre D., Temple G.F., Lorincz A.T.;
RT "Nucleotide sequence of human papillomavirus type 31: a cervical
neoplasia-associated virus."
RL Virology 171:306-311(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04353; AAA46949.1; -
DR PIR; E32444; W4WL31.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.
SQ SEQUENCE 102 AA; 11284 MW; 04E3C9E8ABC5CA6C CRC64;

Query Match 61.5%; Score 56; DB 1; Length 102;
Best Local Similarity 81.8%; Pred. No. 0.49;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 IPKPSPWAPKX 13
  |||||:|||||
Db 32 IPKPAPWAPVK 42

RESULT 5
RIM1_RAT
ID -RIM1 RAT STANDARD; PRT; 1615 AA.
AC Q9JIR4; Q35168;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Regulating synaptic membrane exocytosis protein 1 (Rab3-interacting
molecule 1) (RIM 1).
GN RIMS1 OR RIM1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RAB3A AND RAB3C.
RC TISSUE=Brain;
RX MEDLINE=97394473; PubMed=9252191;
RA Wang Y., Okamoto M., Schmitz F., Hofmann K., Suedhof T.C.;
```


RT "Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion.";
 RL Nature 388:593-598(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1), ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=20347919; PubMed=1048113;
 RA Wang Y., Sugita S., Suedhof T.C.;
 RT "The RIM/NIM family of neuronal C2 domain proteins. Interactions with Rab3 and a new class of Src homology 3 domain proteins.";
 RL J. Biol. Chem. 275:20033-20044(2000).
 CC -!- FUNCTION: Rab effector involved in exocytosis. May act as scaffold protein.
 CC -!- SUBUNIT: Binds SNAP25, SYTI and CACNA1B. Interaction with SYTI is enhanced by calcium ions. Interaction with SNAP25 is weaker in the presence of calcium ions. Binds RAB3A, RAB3B and RAB3D that have been activated by GTP-binding. Binds UNC13 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein associated with plasma membranes from synaptic junctions. Not detected in synaptic vesicles. Detected in presynaptic nerve terminals close to the active zone. Detected in synaptic ribbons of ribbon synapses of retinal photoreceptor cells.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=Rim1B;
 CC IsoId=Q9JIR4-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9JIR4-2; Sequence=VSP_008172;
 CC -!- TISSUE SPECIFICITY: Highly expressed in hippocampus, brain cortex, cerebellum and olfactory bulb. Detected at lower levels in midbrain, hindbrain and spinal cord. Detected retina and in spinal cord motor neurons.
 CC -!- SIMILARITY: Contains 2 C2 domains.
 CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -!- SIMILARITY: Contains 1 Rab-binding (RABBD) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF007836; AAB66703.1; -;
 DR EMBL; AF199333; AAF81655.1; -;
 DR PIR; T03301; T03301.
 DR HGSP; P21707; IRSY.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR008973; C2_CaLB.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR003315; RPH3A effector.
 DR InterPro; IPR000306; Znf_FYVE.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF02318; RPH3A effector; 1.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50004; C2_DOMAIN_2; 2.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS50916; RABBD; 1.
 DR PROSITE; PS50178; ZF_FYVE; 1.
 KW Metal-binding; Zinc; Repeat; Zinc-finger; Alternative splicing.
 FT DOMAIN 22 205
 FT ZN FING 133 193
 FT DOMAIN 422 455
 FT DOMAIN 444 483
 FT DOMAIN 619 705
 FT DOMAIN 758 864
 FT DOMAIN 1259 1301
 FT DOMAIN 1461 1563

FT VARSPLIC 1107 1168 Missing (in isoform 2).
 FT FTIC=VSP_008172.
 SQ SEQUENCE 1615 AA; 179654 MW; 80E76F74BF35FB7E CRC64;
 Query Match 59.3%; Score 54; DB 1; Length 1615;
 Best Local Similarity 53.8%; Pred. No. 12;
 Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
 QY 2 PIPKSPWAPKKH 14
 Db 890 ELPQSPFMPRRH 902
 :|:|:|:|:|:
 :|:|:|:|:|:
 RESULT 6
 RIM1 HUMAN
 ID RIM1 HUMAN STANDARD; PRT; 1692 AA.
 AC Q85UR5; O15048; Q8TDY9; Q8TDZ5; Q9HBA1; Q9HBA2; Q9HBA3; Q9HBA4;
 AC Q9HBA5; Q9HBA6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Regulating synaptic membrane exocytosis protein 1 (Rab3-interacting molecule 1) (RIM 1).
 DE RIMS1 OR RIM1 OR KIAA0340.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), ALTERNATIVE SPLICING, AND VARIANT CORD7 HIS-820.
 RX MEDLINE=22546706; PubMed=12659814;
 RA Johnson S., Halford S., Morris A.G., Patel R.J., Wilkie S.E., Hardcastle A.J., Moore A.T., Zhang K., Hunt D.M.;
 RA "Genomic organisation and alternative splicing of human RIM1, a gene implicated in autosomal dominant cone-rod dystrophy (CORD7).";
 RL Genomics 81:304-314(2003).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC TISSUE=Neuroblastoma;
 RA Aoyama M., Asai K., Shishikura T., Ohira M., Inuzuka H., Morohashi A., Kato T., Nakagawa A.;
 RA "Identification of the alternative form of human RIM.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 RN [4]
 RP SEQUENCE OF 776-1692 FROM N.A. (ISOFORMS 3; 4; 5; 6; 7 AND 8), MUTAGENESIS OF 796-LYS-LYS-797 AND 1591-ARG-ARG-1592, AND INTERACTION WITH SNAP25; SYTI AND CACNA1B.
 RC TISSUE=Brain;
 RX MEDLINE=21413925; PubMed=11438518;
 RA Coppola T., Magnin-Luethi S., Perret-Menoud V., Gattesco S., Schiavo G., Regazzi R.;
 RA "Direct interaction of the Rab3 effector RIM with Ca2+ channels, SNAP-25, and synaptotagmin.";
 RL J. Biol. Chem. 276:32756-32762(2001).
 CC -!- FUNCTION: Rab effector involved in exocytosis. May act as scaffold protein.
 CC -!- SUBUNIT: Binds RAB3A, RAB3B and RAB3D that have been activated by GTP-binding. Binds UNC13 (By similarity). Binds SNAP25, SYTI and CACNA1B. Interaction with SYTI is enhanced by calcium ions.
 CC Interaction with SNAP25 is weaker in the presence of calcium ions.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=8;

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL; L42133; AAB39897.1; -;
 DR EMBL; Z50046; CAA90373.1; -;
 DR PIR; S57977; S57977.
 DR GerMOnline; 140642; -;
 DR SGD; S0002558; CTH1.
 DR InterPro; IPR000571; Znf_CCCH.
 DR Pfam; PF00642; zf-CCCH; 2.
 DR SMART; SM00356; Znf_C3H1; 2.
 DR Nuclear protein; Repeat; Metal-binding; Zinc-finger; DNA-binding.
 KW ZN_FING 210 229 C3H1-TYPE 1.
 FT ZN_FING 248 267 C3H1-TYPE 2.
 FT CONFLICT 142 143 RV -> EI (IN REF. 2).
 SQ SEQUENCE 325 AA; 36785 MW; FID3F46F8BF89DB CRC64;

Query Match 53.8%; Score 49; DB 1; Length 325;
 Best Local Similarity 70.0%; Pred. No. 12;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 4 PKPSFWAPKK 13
 Db 79 PRPSFWLPKK 88
 |:|||||
 79 PRPSFWLPKK 88

RESULT 8
 ID GLRF HUMAN STANDARD; PRT; 1513 AA.
 AC Q9NRV4; Q14452; Q9COE1;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Glucocorticoid receptor DNA binding factor 1 (Glucocorticoid receptor
 DE repression factor 1) (GRF-1) (Rho GAP p190A) (p190-A).
 GN GRF1 OR GRF1 OR KIAA1722.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Breast cancer;
 RX MEDLINE=20510021; PubMed=11054565;
 RA Tikoo A., Czekay S., Viars C., White S., Heath J.K., Arden K.,
 RA Maruta H.;
 RT "p190-A, a human tumor suppressor gene, maps to the chromosomal region
 RT 19q13.3 that is reportedly deleted in some gliomas."
 RL Gene 257:23-31(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=21082932; PubMed=11214970;
 RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 7:347-355(2000).
 RN [3]
 RP SEQUENCE OF 377-1453 FROM N.A. (ISOFORM 1), AND FUNCTION.
 RC TISSUE=Breast cancer;
 RX MEDLINE=91373352; PubMed=1894621;
 RA LeClerc S., Palaniwami R., Xie B.X., Govindan M.V.;
 RT "Molecular cloning and characterization of a factor that binds the
 RT human glucocorticoid receptor gene and represses its expression."
 RL J. Biol. Chem. 266:17333-17340(1991).
 CC -!- FUNCTION: Represses transcription of the glucocorticoid receptor

CC by binding to the cis-acting regulatory sequence 5'-
 CC GAGAAAGAACTGGAGAACAC-3'. May participate in the regulation of
 CC retinal development and degeneration. May transduce signals from
 CC p21-ras to the nucleus, activating via the ras GTP-ase activating
 CC protein (GAP). May also act as a tumor suppressor.
 CC -!- SUBUNIT: Interacts with p120GAP.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9NRV4-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9NRV4-2; Sequence=VSP_007105;
 CC Note=No experimental confirmation available;
 CC -!- PTM: Tyrosine phosphorylated (By similarity).
 CC -!- SIMILARITY: Contains 1 Rho-GAP domain.
 CC -!- SIMILARITY: Contains 4 FF domains.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in positions 533; 540; 607 and 614.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
 CC in positions 389, 533, 540; 607, 614; 1167, 1241, 1292, 1334 and
 CC 1446.
 CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----

EMBL; AF159851; AAF80386.1; ALT_FRAME.
 DR EMBL; AB051509; BAB21813.2; -;
 DR EMBL; M73077; AAA58618.1; ALT_FRAME.
 DR TRANSFAC; T00940; -;
 DR Genew; HGNC:4591; GRF1.
 DR MIM; 605277; -;
 DR GO; GO:0005634; C:nucleus; IC.
 DR GO; GO:0016564; P:transcriptional repressor activity; IDA.
 DR GO; GO:0016481; P:negative regulation of transcription; IDA.
 DR GO; GO:0000074; P:regulation of cell cycle; NAS.
 DR GO; GO:0042478; P:regulation of eye photoreceptor development; ISS.
 DR InterPro; IPR002713; FF.
 DR InterPro; IPR001806; Ras trnsfrmg.
 DR InterPro; IPR008936; Rho GAP.
 DR InterPro; IPR000198; RhoGAP.
 DR Pfam; PF01846; FF; 2.
 DR PRINTS; PR00620; RhoGAP; 1.
 DR PRINTS; PR00449; RASTNSFRMG.
 DR SMART; SM00441; FF; 4.
 DR SMART; SM00324; RhoGAP; 1.
 DR PROSITE; PS02338; RHO GAP; 1.
 DR GTPase activation; DNA-binding; Repressor; Transcription regulation;
 KW Anti-oncogene; Nuclear protein; Phosphorylation; Repeat;
 KW Alternative splicing.
 FT DOMAIN 270 327 FF 1.
 FT DOMAIN 369 422 FF 2.
 FT DOMAIN 429 483 FF 3.
 FT DOMAIN 485 539 FF 4.
 FT DOMAIN 1249 1436 RHO-GAP.
 FT DOMAIN 1440 1507 PRO-RICH.
 FT VARSPLIC 1492 1513 PLPHRPTQPSRNIPQWTKXH -> QLOAHTL (in
 FT isoform 2).
 FT /FTid=VSP_007105.
 FT R -> P (IN REF. 1 AND 3).
 FT V -> G (IN REF. 1 AND 3).
 FT W -> M (IN REF. 1 AND 3).
 FT Q -> A (IN REF. 1 AND 3).
 FT M -> T (IN REF. 1 AND 3).
 FT C -> S (IN REF. 1).
 FT M -> I (IN REF. 1).
 FT PS -> RN (IN REF. 3).
 FT CONFLICT 251 251
 FT CONFLICT 309 309
 FT CONFLICT 362 362
 FT CONFLICT 388 388
 FT CONFLICT 414 414
 FT CONFLICT 474 474
 FT CONFLICT 978 978
 FT CONFLICT 1292 1292
 FT CONFLICT 1452 1452

```

SQ SEQUENCE 1513 AA; 172227 MW; 305C2B3D06519F96 CRC64;
Query Match 53.3%; Score 48.5; DB 1; Length 1513;
Best Local Similarity 52.9%; Pred. NO. 58;
Matches 9; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

QY 1 RPIPKPS---PWAPKKH 14
|||:|||||:|
Db 1497 RPTQPSRNIPQYKEH 1513

RESULT 9
VE4 HPV18
ID -VE4 HPV18 STANDARD; PRT; 88 AA.
AC P06791.
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-JAN-1988 (Rel. 06, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 18.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
CC Papillomavirus.
CX NCBI_TaxID=10582;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87283882; PubMed=3039146;
RA Cole S.T., Danos O.;
RT "Nucleotide sequence and comparative analysis of the human
RT papillomavirus type 18 genome. Phylogeny of papillomaviruses and
RT repeated structure of the B6 and E7 gene products.";
RL J. Mol. Biol. 193:599-608(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05015; CAA28668.1; -.
DR PIR; E26251; W4WL18.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_B4; 1.
KW Early protein.
SQ SEQUENCE 88 AA; 9857 MW; D05F6200BF367B10 CRC64;

Query Match 52.7%; Score 48; DB 1; Length 88;
Best Local Similarity 53.6%; Pred. NO. 4.7;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 IPKPSPWAPKK 13
|||:|||||:|
Db 28 IPAPCFWAPQR 38

RESULT 10
RIM2 MOUSE
ID -RIM2 MOUSE STANDARD; PRT; 1530 AA.
AC Q9EQZ7; Q8CCK2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Regulating synaptic membrane exocytosis protein 2 (Rab3-interacting
DE molecule 2) (RIM 2) (Rab3 interacting protein 2).
GN RIMS2 OR RIM2 OR RAB3IP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), TISSUE SPECIFICITY, AND INTERACTION
WITH RAB3A AND EPAC2.

```

```

RC TISSUE=Insulinoma;
RX MEDLINE=20512528; PubMed=11056535;
RA Ozaki N., Shibasaki T., Kashima Y., Miki T., Takahashi K., Ueno H.,
RA Sunaga Y., Yano H., Matsura Y., Iwanaga T., Takai Y., Seino S.;
RT "cAMP-GEFII is a direct target of cAMP in regulated exocytosis.";
RL Nat. Cell Biol. 2:805-811(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RC STRAIN=C57BL/6J; TISSUE=Hippocampus, and Olfactory bulb;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sadelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wallestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yangisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Ishii Y., Itoh M., Kagawa I.,
RA Hara A., Hashizume W., Imotani K., Shibata K., Shinagawa A.,
RA Miyazaki A., Sakai K., Sasaki D., Lander E.S., Rogers J.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [3]
RP INTERACTION WITH PCLO.
RX MEDLINE=22384373; PubMed=12401793;
RA Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,
RA Sasaki T., Tajima N., Iwanaga T., Seino S.;
RT "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of
RT cAMP-GEFII, Rim2 piccolo complex in cAMP-dependent exocytosis.";
RL J. Biol. Chem. 277:50497-50502(2002).
CC -!- FUNCTION: Rab effector involved in exocytosis. May act as scaffold
CC protein.
CC -!- SUBUNIT: Binds RAB3A and RAB3B that have been activated by GTP-
CC binding (By similarity). Heterodimer with PCLO. Part of a ternary
CC complex involving PCLO and EPAC2.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q9EQZ7-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9EQZ7-2; Sequence=VSP_008181, VSP_008184;
CC Note=No experimental confirmation available;
CC Name=3;
CC IsoId=Q9EQZ7-3; Sequence=VSP_008182, VSP_008183;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Detected in testis, pituitary and an
CC insulinoma cell line. Detected at low levels in cerebellar cortex.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
CC -!- SIMILARITY: Contains 1 PDZ/DEH domain.
CC -!- SIMILARITY: Contains 1 Rab-binding (RabBD) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 PKPSPWAPKKH 14

Db 328 PKPSEWVKKH 338

RESULT 13

CT81_HUMAN

ID CT81_HUMAN STANDARD; PRT; 454 AA.

AC Q9HIQ7; Q96CG7; Q9HIQ6; Q9H6D1;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Protein C20orf81.

GN C20ORF81

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RA Ishibashi T., Kanehori K., Yoshida M., Watanabe S., Ishida S., Ono Y.,

RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,

RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,

RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,

RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,

RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,

RA Wadatsuna M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,

RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,

RA "NEDO human cDNA sequencing project."

RT Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21638749; PubMed=11780052;

RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,

RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,

RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,

RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby R.R.,

RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,

RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,

RA Hammon E., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,

RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,

RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,

RA Lehaeslatho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,

RA Marsh V.L., Martin S.I., McConachie L.J., McLay K., McMurray A.A.,

RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,

RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,

RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,

RA Skuce C.D., Smith M.I., Soderlund C., Steward C.A., Sulston J.E.,

RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,

RA Tracey A., Tromans A.C., Vaubin M., Wall M., Wallis J.M.,

RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,

RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,

RA Rogers J.;

RT "The DNA sequence and comparative analysis of human chromosome 20.";

RL Nature 414:865-871 (2001).

RN [3]

RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).

RC TISSUE=Blood, Brain, and Skin;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,

RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Frange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=Q9HIQ7-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9HIQ7-2; Sequence=VSP_003818;
CC Name=3;
CC IsoId=Q9HIQ7-3; Sequence=VSP_003819;
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AK056353; BABY1160.1; -;
DR EMBL; AK026029; BABY15328.1; -;
DR EMBL; AL161656; CAC21463.1; -;
DR EMBL; AL161656; CAC21464.1; -;
DR EMBL; BC014247; AAH14247.1; -;
DR EMBL; BC014247; AAH14247.1; -;
DR EMBL; BC037240; AAH37240.2; -;
DR EMBL; BC051816; AAH51816.1; -;
DR EMBL; HGNC:16212; C20orf81.
KW Alternative splicing.

FT DOMAIN 232 264 HIS-RICH.
FT DOMAIN 312 315 POLY-PRO.
FT VARSPLIC 69 119 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 264 354 Missing (in isoform 3).
FT CONFLICT 253 253 A -> V (IN REF. 3; AAH14247).
FT CONFLICT 430 430 R -> K (IN REF. 1; BAB15328).
SQ SEQUENCE 454 AA; 51782 MW; 6D783D66F3FD779A CRC64;

Query Match Best Local Similarity 50.5%; Score 46; DB 1; Length 454;
Matches 9; Conservative 1; Mismatches 4; Indels 4; Gaps 1;

QY 2 PIPKSP-----WAPKKHR 15

Db 388 PIPGPNPHQHWGPVVRH 405

RESULT 14

SYGB_HAEIN

ID SYGB_HAEIN STANDARD; PRT; 688 AA.

AC P43822;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
DE beta chain) (GLYRS).

GN GLYS OR H10924.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI_TaxID=727;

RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
EX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC + glycyI-tRNA(Gly).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32774; AAC22584.1; -.
CC PIR; B64103; B64103.
CC TIGR; HI0924; -.
CC HAWAP; MF_00255; -.
CC InterPro; IPR002311; tRNA_synt_2f.
CC InterPro; IPR006194; tRNA_synt_Gly.
CC Pfam; PF02092; tRNA_synt_2f; 1.
CC PRINTS; PR01045; TRNASYNTHGB.
CC TIGRFAMs; TIGR00211; GlyS; 1.
CC PROSITE; PS50861; AA TRNA_LIGASE_II GLYAB; 1.
CC Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
CC SEQUENCE 688 AA; 75612 MW; 67F9ACEF00A94184 CRC64;
DR DR
DR PIR; B64103; B64103.
DR TIGR; HI0924; -.
DR HAWAP; MF_00255; -.
DR InterPro; IPR002311; tRNA_synt_2f.
DR InterPro; IPR006194; tRNA_synt_Gly.
DR Pfam; PF02092; tRNA_synt_2f; 1.
DR PRINTS; PR01045; TRNASYNTHGB.
DR TIGRFAMs; TIGR00211; GlyS; 1.
DR PROSITE; PS50861; AA TRNA_LIGASE_II GLYAB; 1.
DR Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 688 AA; 75612 MW; 67F9ACEF00A94184 CRC64;

Query Match 50.5%; Score 46; DB 1; Length 688;
Best Local Similarity 72.7%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PIPKPSWPAPK 12
Db 143 PIPKPMWADK 153
|||||
|

RESULT 15
SYGB_PASMU STANDARD; PRT; 689 AA.
AC P57905;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
DE beta chain) (GlyRS).
DE GLYS OR PM102.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OC NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
```

```

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate
CC + glycyI-tRNA(Gly).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006151; AAK03186.1; -.
CC HAWAP; MF_00255; -.
CC InterPro; IPR002311; tRNA_synt_2f.
CC InterPro; IPR006194; tRNA_synt_Gly.
CC Pfam; PF02092; tRNA_synt_2f; 1.
CC PRINTS; PR01045; TRNASYNTHGB.
CC TIGRFAMs; TIGR00211; GlyS; 1.
CC PROSITE; PS50861; AA TRNA_LIGASE_II GLYAB; 1.
CC Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 689 AA; 76019 MW; D67A980B5143B21E CRC64;

Query Match 50.5%; Score 46; DB 1; Length 689;
Best Local Similarity 72.7%; Pred. No. 59;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PIPKPSWPAPK 12
Db 143 PIPKPMWADK 153
|||||
|

Search completed: May 27, 2004, 16:19:03
Job time : 7.92308 secs
```



```

DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE E4 protein (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPV16E2CC4;
RA Watts K.J., Thompson C.H., Cossart Y.E., Rose B.R.;
RT "Sequence variation and physical state of human Papillomavirus type 16
RT cervical cancer isolates from Australia and New Caledonia.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF407217; AAL01395.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON_TER 1
SQ SEQUENCE 95 AA; 10608 MW; AED4269D05E307CE CRC64;

Query Match 100.0%; Score 91; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIKPSPWAPKKHR 15
Db 29 RPIKPSPWAPKKHR 43

RESULT 3
Q8B5P2
ID Q918T7 PRELIMINARY; PRT; 95 AA.
AC Q918T7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE E4 protein (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPV16E2CC4;
RA Watts K.J., Thompson C.H., Cossart Y.E., Rose B.R.;
RT "Sequence variation and physical state of human Papillomavirus type 16
RT cervical cancer isolates from Australia and New Caledonia.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF407218; AAL01399.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON_TER 1
SQ SEQUENCE 95 AA; 10580 MW; BE0E469D177307CE CRC64;

Query Match 100.0%; Score 91; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIKPSPWAPKKHR 15
Db 29 RPIKPSPWAPKKHR 43

RESULT 4
Q8B5P2
ID Q8B5P2 PRELIMINARY; PRT; 95 AA.
AC Q8B5P2;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE E4 protein (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPV16E2CC4;
RA Watts K.J., Thompson C.H., Cossart Y.E., Rose B.R.;
RT "Sequence variation and physical state of human Papillomavirus type 16
RT cervical cancer isolates from Australia and New Caledonia.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF407218; AAL01399.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON_TER 1
SQ SEQUENCE 95 AA; 10580 MW; BE0E469D177307CE CRC64;

Query Match 100.0%; Score 91; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIKPSPWAPKKHR 15
Db 29 RPIKPSPWAPKKHR 43

RESULT 5
Q8B5N7
ID Q8B5N7 PRELIMINARY; PRT; 95 AA.
AC Q8B5N7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Putative E4 protein (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RA Terai M., Ma Z., Burk R.D.;
RT "Cloning and Sequencing of Non-European Human Papillomavirus (HPV)
RT Variants Complete Genomes from Patients with Cervical Cancer by an
RT Overlapping PCR Method.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF472508; AAO15709.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON_TER 1
SQ SEQUENCE 95 AA; 10542 MW; AED17903867307CE CRC64;

Query Match 100.0%; Score 91; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIKPSPWAPKKHR 15
Db 29 RPIKPSPWAPKKHR 43

RESULT 6
Q80MM0
ID Q80MM0 PRELIMINARY; PRT; 95 AA.
AC Q80MM0;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE E4 (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RA Terai M., Ma Z., Burk R.D.;
RT "Cloning and Sequencing of Non-European Human Papillomavirus (HPV)
RT Variants Complete Genomes from Patients with Cervical Cancer by an
RT Overlapping PCR Method.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF472509; AAO15709.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON_TER 1
SQ SEQUENCE 95 AA; 10542 MW; AED17903867307CE CRC64;

Query Match 100.0%; Score 91; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIKPSPWAPKKHR 15
Db 29 RPIKPSPWAPKKHR 43

```

```

DE Putative E4 protein (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RA Terai M., Ma Z., Burk R.D.;
RT "Cloning and Sequencing of Non-European Human Papillomavirus (HPV)
RT Variants Complete Genomes from Patients with Cervical Cancer by an
RT Overlapping PCR Method.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF472508; AAO15701.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON_TER 1
SQ SEQUENCE 95 AA; 10585 MW; AEC1961D677307CE CRC64;

Query Match 100.0%; Score 91; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIKPSPWAPKKHR 15
Db 29 RPIKPSPWAPKKHR 43

RESULT 5
Q8B5N7
ID Q8B5N7 PRELIMINARY; PRT; 95 AA.
AC Q8B5N7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Putative E4 protein (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RA Terai M., Ma Z., Burk R.D.;
RT "Cloning and Sequencing of Non-European Human Papillomavirus (HPV)
RT Variants Complete Genomes from Patients with Cervical Cancer by an
RT Overlapping PCR Method.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF472509; AAO15709.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON_TER 1
SQ SEQUENCE 95 AA; 10542 MW; AED17903867307CE CRC64;

Query Match 100.0%; Score 91; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIKPSPWAPKKHR 15
Db 29 RPIKPSPWAPKKHR 43

RESULT 6
Q80MM0
ID Q80MM0 PRELIMINARY; PRT; 95 AA.
AC Q80MM0;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE E4 (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RA Terai M., Ma Z., Burk R.D.;
RT "Cloning and Sequencing of Non-European Human Papillomavirus (HPV)
RT Variants Complete Genomes from Patients with Cervical Cancer by an
RT Overlapping PCR Method.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF472509; AAO15709.1; -.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON_TER 1
SQ SEQUENCE 95 AA; 10542 MW; AED17903867307CE CRC64;

Query Match 100.0%; Score 91; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 6.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPIKPSPWAPKKHR 15
Db 29 RPIKPSPWAPKKHR 43

```

RP	SEQUENCE
RC	STRAIT

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OX Alphavirus.
OX NCBI_TaxID=11021;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=BR56-BeAn122;
RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159559; AAF04801.1; -.
DR HSSP; P03315; 1VCP.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000336; Flavi_glycoprote.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR000930; Peptidase_S3.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
DR Polyprotein.
KW SEQUENCE 1242 AA; 137436 MW; 25A0990E69D16636 CRC64;
QY Query Match 61.5%; Score 56; DB 12; Length 1242;
Best Local Similarity 66.7%; Pred. No. 9;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Db 1 RPIPKPSWAPKKR 15
:|||||
70 KPAPKPKPAAPKKR 84
RESULT 11
Q9PZW7 PRELIMINARY; PRT; 1242 AA.
ID Q9PZW7
AC Q9PZW7
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Structural polyprotein.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
OS virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OX Alphavirus.
OX NCBI_TaxID=11021;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=PA86-435731;
RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159560; AAF04802.1; -.
DR HSSP; P03315; 1VCP.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR Polyprotein.
KW SEQUENCE 1242 AA; 137405 MW; 7254B6D64DED65C3 CRC64;
QY Query Match 61.5%; Score 56; DB 12; Length 1242;
Best Local Similarity 66.7%; Pred. No. 9;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Db 1 RPIPKPSWAPKKR 15
:|||||
70 KPAPKPKPAAPKKR 84
RESULT 11
Q9PZW7 PRELIMINARY; PRT; 1242 AA.
ID Q9PZW7
AC Q9PZW7
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Structural polyprotein.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
OS virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OX Alphavirus.
OX NCBI_TaxID=11021;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=BR85-436087;
RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159561; AAF04803.1; -.
DR HSSP; P03315; 1VCP.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000336; Flavi_glycoprote.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000930; Peptidase_S3.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
DR Polyprotein.
KW SEQUENCE 1242 AA; 137344 MW; 2759F20F690B4A13 CRC64;
QY Query Match 61.5%; Score 56; DB 12; Length 1242;
Best Local Similarity 66.7%; Pred. No. 9;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Db 1 RPIPKPSWAPKKR 15
:|||||
70 KPAPKPKPAAPKKR 84
RESULT 12
Q9PZW6 PRELIMINARY; PRT; 1242 AA.
ID Q9PZW6
AC Q9PZW6
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Structural polyprotein.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
OS virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OX Alphavirus.
OX NCBI_TaxID=11021;
RN [1]_SEQUENCE FROM N.A.
RP STRAIN=BR85-436087;
RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159561; AAF04803.1; -.
DR HSSP; P03315; 1VCP.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0004252; F:serine-type endopeptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000336; Flavi_glycoprote.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000930; Peptidase_S3.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
DR Polyprotein.
KW SEQUENCE 1242 AA; 137405 MW; 7254B6D64DED65C3 CRC64;
QY Query Match 61.5%; Score 56; DB 12; Length 1242;
Best Local Similarity 66.7%; Pred. No. 9;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Db 1 RPIPKPSWAPKKR 15
:|||||
70 KPAPKPKPAAPKKR 84

Best Local Similarity 58.8%; Pred. No. 5.1;
Matches 10; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

Qy 2 PIPKPSFW----APKKH 14
| | | | | | | | | |
Db 139 PIPKPSFWRTAPTSPH 155

RESULT 15
Q95W83 PRELIMINARY; PRT; 2527 AA.
ID Q95W83
AC Q95W83;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VARI.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21442075; PubMed=11557894;
RA Flick K., Scholander C., Chen Q., Fernandez V., Pouvelle B., Gysin J.,
RA Wahlgren M.;
RT "Role of nonimmune IgG bound to PfEMP1 in placental malaria.";
RL Science 293:2098-2100(2001).
DR EMBL; AF366567; AAL12845.1; -;
DR GO; GO:0005539; F:glycosaminoglycan binding; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR InterPro; IPR004258; PfEMP.
DR Pfam; PF03011; PfEMP; 2.
FT NON_TER 2527 2527
SQ SEQUENCE 2527 AA; 292137 MW; 59BE2766BF0425E4 CRC64;

Query Match 58.2%; Score 53; DB 5; Length 2527;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 RPIKPSFWAPKKHR 15
: | | | | | | | | | |
Db 2401 KEVKPGPKAPKKR 2415

Search completed: May 27, 2004, 16:20:56
Job time : 29.8462 secs

RESULT 13
O49870 PRELIMINARY; PRT; 330 AA.
ID O49870
AC O49870;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Extensin (Fragment).
GN EX1.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RA Doan D.N.P., Sturaro M., Olsen O.A.;
RT "Characterization of a nuclear cDNA encoding a putative extensin from
RT developing barley grains (Hordeum vulgare L.).";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98204; CAB10894.1; -;
DR PIR; T05717; T05717. 1
FT NON_TER 1
SQ SEQUENCE 330 AA; 34218 MW; E441C3216470A749 CRC64;

Query Match 59.3%; Score 54; DB 10; Length 330;
Best Local Similarity 75.0%; Pred. No. 4.7;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RPIKPSFWAPK 12
: | | | | | | | | | |
Db 200 KEVKPSFWAPK 211

RESULT 14
Q9LIP1 PRELIMINARY; PRT; 356 AA.
ID Q9LIP1
AC Q9LIP1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Gb|AAD31058.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL; AP001298; BAB02195.1; -;
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR InterPro; IPR001865; Ribosomal S2.
DR PROSITE; PS00962; RIBOSOMAL_S2_1; 1.
SQ SEQUENCE 356 AA; 39906 MW; 124AB52ADE8C08EF CRC64;

Query Match 59.3%; Score 54; DB 10; Length 356;

Result	No.	Query %			DB	ID	Description
		Score	Match	Length			
1	75	100.0	12	2	AAW62279	AAW62279 HPV16 hyd	
2	75	100.0	15	2	AAW62278	AAW62278 HPV16 hyd	
3	75	100.0	20	2	AAR15562	AAR15562 Immunopep	
4	75	100.0	26	2	AAW62277	AAW62277 HPV16 hyd	
5	75	100.0	26	5	AAU75260	AAU75260 Hydrophil	
6	75	100.0	30	2	AAR15564	AAR15564 Immunopep	
7	68	90.7	14	2	AAR14288	AAR14288 Seroreact	
8	62	82.7	11	2	AAR14287	AAR14287 Seroreact	
9	56	74.7	105	4	AAO03023	AAO03023 Human pol	
10	50	66.7	55	4	ABG16750	ABG16750 Novel hum	
11	50	66.7	115	4	AAO04150	AAO04150 Human pol	
12	46	61.3	72	4	ABG08584	ABG08584 Novel hum	
13	46	61.3	91	6	ABR41361	ABR41361 Human DIT	
14	46	61.3	121	3	AAQ01947	AAQ01947 Human sec	
15	46	61.3	230	5	ABU05501	ABU05501 M. tuberc	
16	46	61.3	394	4	ABG16932	ABG16932 Novel hum	
17	46	61.3	1255	4	ABBS58580	ABBS58580 Drosophil	
18	45	60.0	73	4	AAU40297	AAU40297 Protonib	
19	45	60.0	73	6	ABM36816	ABM36816 Protonib	
20	45	60.0	138	4	AAO00838	AAO00838 Human pol	
21	45	60.0	312	6	ABP75923	ABP75923 Human sec	
22	45	60.0	1053	7	ADES4638	ADES4638 Human Fro	
23	45	60.0	1553	7	ADES4636	ADES4636 Rat Prote	
24	44.5	59.3	1243	4	ABG21281	ABG21281 Novel hum	
25	44.5	59.3	1312	5	ABP69782	ABP69782 Human pol	

CC humans, since over 90% of cervical carcinoma patients show cervical HPV
 CC infection. It is also useful to determine the type(s) of HPV infection in
 CC a patient, by using a molecule binding specifically to a subset of HPV E4
 CC proteins. This is important, since progression to malignant disease (and
 CC hence clinical prognosis) is dependent on HPV type. Molecules capable of
 CC binding E4 are also useful to target anticancer/antiviral agents capable
 CC of destroying papilloma viruses and/or papilloma virus-infected cells.
 CC The present sequence represents a specifically claimed HPV16 peptide
 CC found in the hydrophilic region
 CC
 CC Sequence 12 AA;
 SQ

Query Match 100.0%; Score 75; DB 2; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.00031;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 1 PKPSPWAPKKHR 12

RESULT 2

AAW62278
 ID AAW62278 standard; peptide; 15 AA.

XX
 AC AAW62278;

XX
 DT 24-SEP-1998 (first entry)

XX
 DE HPV16 hydrophilic region peptide #2.

XX
 KW HPV16; human papilloma virus; epithelial tumour; cervical cancer;
 KW precancerous cervical lesion; screening; detection; infection; cervix;
 KW HPV E4.

XX
 OS Human papillomavirus.

XX
 FN WO9825145-A1.

XX
 PD 11-JUN-1998.

XX
 PF 03-DEC-1997; 97WO-GB003321.

XX
 PR 03-DEC-1996; 96GB-00025142.

XX
 PR 05-SEP-1997; 97GB-00018745.

XX
 PA (MEDI-) MEDICAL RES COUNCIL.

XX
 PI Doorbar J;

XX
 DR WPI; 1998-333497/29.

XX
 PT Detecting papilloma virus infection using molecule binding to E4 protein
 PT - useful, e.g. in screening for pre-cancerous cervical lesions and to
 PT determine type(s) of human papilloma virus infecting human patients.

XX
 PS Claim 10; Page 37; 52pp; English.

XX
 CC A new method has been developed for detecting a papilloma virus infection
 CC in an organism. The method comprises: (i) obtaining a sample of cells
 CC from the potential infection site; (ii) contacting the cells with a
 CC molecule binding specifically to papilloma virus E4 protein, and (iii)
 CC monitoring the binding. The method is useful to detect papilloma virus
 CC infections in organisms (especially mammals) and especially HPV
 CC infections (e.g. with HPV16, 18, 33, 35, 45, 51, 56, 58 or 61) in humans.
 CC Papilloma viruses cause epithelial tumours in humans varying in severity
 CC depending on the infection site and HPV type involved. The method is
 CC particularly useful to determine papilloma infection in the mammalian
 CC cervix and especially to screen for pre-cancerous cervical lesions in
 CC humans, since over 90% of cervical carcinoma patients show cervical HPV
 CC infection. It is also useful to determine the type(s) of HPV infection in
 CC a patient, by using a molecule binding specifically to a subset of HPV E4
 CC proteins. This is important, since progression to malignant disease (and

CC hence clinical prognosis) is dependent on HPV type. Molecules capable of
 CC binding E4 are also useful to target anticancer/antiviral agents capable
 CC of destroying papilloma viruses and/or papilloma virus-infected cells.
 CC The present sequence represents a specifically claimed HPV16 peptide
 CC found in the hydrophilic region
 CC
 CC Sequence 15 AA;
 SQ

Query Match 100.0%; Score 75; DB 2; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.00039;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 4 PKPSPWAPKKHR 15

RESULT 3

AAAR15562
 ID AAR15562 standard; protein; 20 AA.

XX
 AC AAR15562;

XX
 DT 02-MAR-1992 (first entry)

XX
 DE Immunopeptide #2 derived from HPV16 E4 peptide.

XX
 KW cervical cancer; cervical intraepithelial neoplasia; CIN; wart;
 KW squamous cell carcinoma; ELISA; HPV 16.

XX
 OS Synthetic.

XX
 FN WO9118294-A.

XX
 PD 28-NOV-1991.

XX
 PF 11-MAY-1990; 90SE-00001705.

XX
 PR 11-MAY-1990; 90SE-00001705.

XX
 PA (MEDS-) MEDSCAND AB.

XX
 PI Dillner J, Dillner L, Cheng HM;

XX
 DR WPI; 1991-369390/50.

XX
 PT Diagnosis of human papilloma virus infection and PV-carrying tumours -
 PT using synthetic peptide(s) to detect virus specific antigen-antibody
 PT complexes by immunoassay.

XX
 PS Disclosure; Page 38; 72pp; English.

XX
 CC This is one of a large number of peptides which have been synthesised on
 CC the basis of the amino acid sequences for the E2, E4, E7, L1 or L2
 CC proteins of HPV 1, 5, 6, 8, 11, 16, 18, 31 and 33. The selection of
 CC peptide sequences was based on the assumption that an immunoreactive
 CC region might be situated in the same relative region of a protein from
 CC different HPV types. The peptides were used in diagnostic immunoassays to
 CC detect HPV-infection. See AAR15523-R15601
 CC
 CC Sequence 20 AA;

Query Match 100.0%; Score 75; DB 2; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.00052;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 2 PKPSPWAPKKHR 13

RESULT 4

AAW62277

ID AAW62277 standard; peptide; 26 AA.
 AC AAW62277;
 XX
 DT 24-SEP-1998 (first entry)
 DE
 XX HPV16 hydrophilic region peptide #1.
 XX
 KW HPV16; human papilloma virus; epithelial tumour; cervical cancer;
 KW precancerous cervical lesion; screening; detection; infection; cervix;
 KW HPV E4.
 XX
 OS Human papillomavirus.
 PN WO9825145-A1.
 XX
 PD 11-JUN-1998.
 XX
 XX 03-DEC-1997; 97WO-GB003321.
 XX
 PR 03-DEC-1996; 96GB-00025142.
 PR 05-SEP-1997; 97GB-00018745.
 XX
 PA (MEDI-) MEDICAL RES COUNCIL.
 XX
 PI Doorbar J;
 XX
 DR WPI; 1998-333497/29.
 XX
 PT Detecting papilloma virus infection using molecule binding to E4 protein
 PT - useful, e.g. in screening for pre-cancerous cervical lesions and to
 PT determine type(s) of human papilloma virus infecting human patients.
 XX
 PS Claim 9; Page 37; 52pp; English.
 XX
 CC A new method has been developed for detecting a papilloma virus infection
 CC in an organism. The method comprises: (i) obtaining a sample of cells
 CC from the potential infection site; (ii) contacting the cells with a
 CC molecule binding specifically to papilloma virus E4 protein, and (iii)
 CC monitoring the binding. The method is useful to detect papilloma virus
 CC infections in organisms (especially mammals) and especially HPV
 CC infections (e.g. with HPV16, 18, 33, 35, 45, 51, 56, 58 or 61) in humans.
 CC Papilloma viruses cause epithelial tumours in humans varying in severity
 CC depending on the infection site and HPV type involved. The method is
 CC particularly useful to determine papilloma infection in the mammalian
 CC cervix and especially to screen for pre-cancerous cervical lesions in
 CC humans, since over 90% of cervical carcinoma patients show cervical HPV
 CC infection. It is also useful to determine the type(s) of HPV infection in
 CC a patient, by using a molecule binding specifically to a subset of HPV E4
 CC proteins. This is important, since progression to malignant disease (and
 CC hence clinical prognosis) is dependent on HPV type. Molecules capable of
 CC binding E4 are also useful to target anticancer/antiviral agents capable
 CC of destroying papilloma viruses and/or papilloma virus-infected cells.
 CC The present sequence represents a specifically claimed HPV16 peptide
 CC found in the hydrophilic region
 XX
 SQ Sequence 26 AA;
 Query Match 100.0%; Score 75; DB 2; Length 26;
 Best Local Similarity 100.0%; Pred. No. 0.00068;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PKPSPWAPKKHR 12
 DB |||||
 4 PKPSPWAPKKHR 15
 RESULT 5
 AAU75260
 ID AAU75260 standard; peptide; 26 AA.
 XX
 AC AAU75260;
 XX
 DT 02-MAR-1992 (first entry)
 XX
 DE Immunopeptide #4 derived from HPV16 E4 peptide.

21-MAY-2002 (first entry)
 Hydrophilic region of HPV16 E4 protein.
 Papilloma virus associated antigen; cell proliferation marker;
 cervical malignancy; human papillomavirus infection; HPV; lesion;
 cellular abnormality; cellular proliferation; cellular growth; dysplasia;
 neoplasia; cancer; papilloma smear test; E4 protein.
 Human papillomavirus type 16.
 WO200208764-A1.
 31-JAN-2002.
 16-MAR-2001; 2001WO-GB001176.
 24-JUL-2000; 2000GB-00018140.
 (MEDI-) MEDICAL RES COUNCIL.
 Doorbar J;
 WPI; 2002-188648/24.
 Detecting abnormalities e.g. abnormal cellular proliferation, in a sample
 from a patient comprises contacting cells with a molecule which can bind
 a papilloma virus associated antigen, or a cell proliferation or viral
 activity marker.
 Disclosure; Page 23; 90pp; English.
 The present invention relates to a method for detecting abnormalities in
 a sample from a patient. The method comprises contacting a sample of the
 patient's cells with two or more molecules, where at least one molecule
 is capable of binding a papilloma virus associated antigen, and at least
 one molecule is capable of binding a cell proliferation marker. The
 method is useful for simultaneously screening for abnormalities which
 indicate or can lead to cervical malignancy, for human papillomavirus
 (HPV) infections, and precursor lesions or other conditions which occur
 with cervical malignancy. The method is also useful for assessing the
 risk associated with cellular abnormality in a patient sample, and for
 determining, assessing or diagnosing the presence or absence of abnormal
 cellular proliferation, cellular growth abnormality, dysplasia,
 neoplasia, or a pre-cancerous or cancerous state in a tissue. The new
 method is much simpler, and yields more information more quickly than
 conventional papilloma smear testing programmes. Compared with previous
 methods of screening, the new method has reduced chances of false
 negatives occurring, requires fewer samples to gain the same amount of
 information, and alleviates the need for repeated or further testing. The
 present sequence represents the hydrophilic region of the HPV16 E4
 protein binds antibody molecules
 Sequence 26 AA;
 Query Match 100.0%; Score 75; DB 5; Length 26;
 Best Local Similarity 100.0%; Pred. No. 0.00068;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 PKPSPWAPKKHR 12
 DB |||||
 4 PKPSPWAPKKHR 15
 RESULT 6
 AAR15564
 ID AAR15564 standard; protein; 30 AA.
 XX
 AC AAR15564;
 XX
 DT 02-MAR-1992 (first entry)
 XX
 DE Immunopeptide #4 derived from HPV16 E4 peptide.

XX cervical cancer; cervical intraepithelial neoplasia; CIN; wart;
 KW squamous cell carcinoma; ELISA; HPV 16.
 XX

OS Synthetic.

XX WO9118294-A.

XX 28-NOV-1991.

XX 11-MAY-1990; 90SE-00001705.

XX 11-MAY-1990; 90SE-00001705.

XX (MEDS-) MEDSCAND AB.

XX Dillner J, Dillner L, Cheng HM;

XX WPI; 1991-369390/50.

XX
 PT Diagnosis of human papilloma virus infection and pv-carrying tumours -
 PT using synthetic peptide(s) to detect virus specific antigen-antibody
 PT complexes by immunoassay.

XX Disclosure; Page 38; 72pp; English.

XX This is one of a large number of peptides which have been synthesised on
 CC the basis of the amino acid sequences for the E2, E4, E7, L1 or L2
 CC proteins of HPV 1, 5, 6, 8, 11, 16, 18, 31 and 33. The selection of
 CC peptide sequences was based on the assumption that an immunoreactive
 CC region might be situated in the same relative region of a protein from
 CC different HPV types. The peptides were used in diagnostic immunoassays to
 CC detect HPV-infection. See AAR15523-R15601

XX Sequence 30 AA;

Query Match 100.0%; Score 75; DB 2; Length 30;
 Best Local Similarity 100.0%; Pred. No. 0.00078;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSFWAPKKHR 12

DB 10 PKPSFWAPKKHR 21

RESULT 7

AAR14288

ID AAR14288 standard; protein; 14 AA.

XX AAR14288;

XX 02-JAN-1992 (first entry)

XX Seroreactive epitope #2 of HPV16 protein E4.

XX HPV16-dependent human disease; E6; E7; L1.

XX Synthetic.

XX EP451550-A.

XX 16-OCT-1991.

XX 19-MAR-1991; 91EP-00104197.

XX 20-MAR-1990; 90EP-00105222.

XX (BEHW) BEHRINGWERKE AG.

XX Muller M, Gissmann L;

XX WPI; 1991-304643/42.

XX

PT Sero-active epitope(s) of human papilloma-virus 16 proteins - for use
 PT vaccines and diagnosis.

XX Claim 1; Page 11; 15pp; English.

XX This is one of two seroreactive epitopes identified from HPV16 protein
 CC E4. Peptides containing this epitope are also claimed and can be used to
 CC generate antibodies to HPV. See AAR14287-R14302 and AAR14168-Q14171

XX Sequence 14 AA;

Query Match 90.7%; Score 68; DB 2; Length 14;

Best Local Similarity 100.0%; Pred. No. 0.0036;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KPSPFWAPKKHR 12

DB 1 KPSPFWAPKKHR 11

RESULT 8

AAR14287

ID AAR14287 standard; peptide; 11 AA.

XX AAR14287;

XX 02-JAN-1992 (first entry)

XX Seroreactive epitope #1 of HPV16 protein E4.

XX HPV16-dependent human disease; E6; E7; L1.

XX Synthetic.

XX EP451550-A.

XX 16-OCT-1991.

XX 19-MAR-1991; 91EP-00104197.

XX 20-MAR-1990; 90EP-00105222.

XX (BEHW) BEHRINGWERKE AG.

XX Muller M, Gissmann L;

XX WPI; 1991-304643/42.

XX Sero-active epitope(s) of human papilloma-virus 16 proteins - for use
 PT vaccines and diagnosis.

XX Claim 1; Page 11; 15pp; English.

XX This is one of two seroreactive epitopes identified from HPV16 protein
 CC E4. Peptides containing this epitope are also claimed and can be used to
 CC generate antibodies to HPV. See also AAR14288-R14302 and AAR14168-Q14171

XX Sequence 11 AA;

Query Match 82.7%; Score 62; DB 2; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.02;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPFWAPKK 10

DB 2 KPSPFWAPKK 11

RESULT 9

AAO03023

ID AAO03023 standard; protein; 105 AA.

XX AAO03023;

```

XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 16915.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US004927.
XX PR 28-FEB-2000; 2000US-00515126.
XX PS 18-MAY-2000; 2000US-00577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX PD WPI; 2001-514838/56.
XX DR N-PSDB; AAI82954.
XX PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX PT and treating e.g. leukemia, inflammation and immune disorders.
XX PS Claim 20; SEQ ID NO 16915; 1399pp + Sequence Listing; English.
XX CC The invention relates to human polynucleotides (AA179941-AA193841) and
XX CC the encoded proteins (AA00010-AA013910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation. Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 105 AA;
Query Match 74.7%; Score 56; DB 4; Length 105;
Best Local Similarity 81.8%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 PKPSPWAPKKH 11
DB 82 PCPSPWAPKTH 92
RESULT 10
ID ABG16750 standard; protein; 55 AA.
XX AC ABG16750;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #16741.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 16915.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PS 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX PD WPI; 2001-639362/73.
XX DR N-PSDB; AAS80937.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 47109; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 55 AA;
Query Match 66.7%; Score 50; DB 4; Length 55;
Best Local Similarity 63.8%; Pred. No. 5.2;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 PKPSPWAPKKH 11
DB 12 PPPPPWAPARH 22
RESULT 11
ID AAO04150 standard; protein; 115 AA.
XX AC AAO04150;
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 18042.
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX

```

PN	WO200164835-A2.	XX	WPI; 2001-639362/73.	XX	The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activities. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX		DR	N-PSDB; AAS72771.	DR	Sequence 72 AA;
XX		XX		XX	Query Match 61.3%; Score 46; DB 4; Length 72;
PT	26-FEB-2001; 2001WO-US004927.	PT	New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.	PT	Best Local Similarity 60.0%; Pred. No. 25;
PR	28-FEB-2000; 2000US-00515126.	PR		PR	Matches 6; Conservative 3; Mismatches
PR	18-MAY-2000; 2000US-00577409.	PR		PR	QY 3 PSPWAPKKHR 12
XX	(HYSE-) HYSEQ INC.	XX		XX	DB 20 PTFWAPVQHK 29
PA	Tang YT, Liu C, Drmanac RT;	PA	Claim 20; SEQ ID NO 38943; 103pp; English.	PA	RESULT 13
PI	WPI; 2001-514838/56.	PI		PI	ABR41361
DR	N-PSDB; AAI84081.	DR		DR	ID ABR41361 standard; protein; 91 AA.
XX		XX		XX	AC ABR41361;
XX		XX		XX	DT 02-JUN-2003 (first entry)
PT	Isolated nucleic acids and polypeptides, useful for preventing diagnosing and treating e.g. leukemia, inflammation and immune disorders.	PT		PT	DE Human DITHP transcription factor.
XX	Claim 20; SEQ ID NO 18042; 1399pp + Sequence Listing; English.	XX		XX	Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis; cancer; cell proliferative disorder; autoimmune disorder;
XX		XX		XX	KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
CC	The invention relates to human polynucleotides (AAI79941-AAI93841) and the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	CC		CC	KW connective tissue disorder; drug screening; proteome analysis;
XX		XX		XX	KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in; disease model; toxicological testing; transcript imaging;
XX		XX		XX	KW transcription factor.
SQ	Sequence 115 AA;	SQ		SQ	OS Homo sapiens.
	Query Match 66.7%; Score 50; DB 4; Length 115;				XX WO200297031-A2.
	Best Local Similarity 72.7%; Pred. No. 11;				XX 05-DEC-2002.
	Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;				PD 27-MAR-2002; 2002WO-US010056.
QY	1 PKPSPWAPKKH 11	QY		QY	XX 28-MAR-2001; 2001US-0279619P.
DB	91 PFPPPWAPKKK 101	DB		DB	PR 29-MAR-2001; 2001US-0280067P.
RESULT 12					
ABG08584					
ID	ABG08584 standard; protein; 72 AA.				
XX					
AC	ABG08584;				
XX					
DT	13-FEB-2002 (first entry)				
DE	Novel human diagnostic protein #8575.				
XX					
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;				
KW	food supplement; medical imaging; diagnostic; genetic disorder.				
XX					
OS	Homo sapiens.				
XX					
PN	WO200175067-A2.				
XX					
PD	11-OCT-2001.				
XX					
PF	30-MAR-2001; 2001WO-US008631.				
XX					
PR	31-MAR-2000; 2000US-00540217.				
PR	23-AUG-2000; 2000US-00649167.				
XX					
XX	(HYSE-) HYSEQ INC.				
PA	Drmanac RT, Liu C, Tang YT;				
PI					

PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
PI Daugherty SC, Dam TC, Liu TP, Nguyen DA, Kleefeld Y, Gerstein EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX
XX WPI; 2003-129518/12.
DR N-PSDE; ACC46301.
XX
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
XX Claim 27; SEQ ID NO 896; 591pp; English.
XX
XX The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a DITHP protein which has transcription
CC factor activity. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 91 AA;
SQ
Query Match 61.3%; Score 46; DB 6; Length 91;
Best Local Similarity 63.6%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 PKPSPWAPKKH 11
Db 22 PPSPPWPPGPH 32
RESULT 14
AAG01947
ID AAG01947 standard; protein; 121 AA.
XX
AC AAG01947;
XX
XX 06-OCT-2000 (first entry)
DT

XX Human secreted protein, SEQ ID NO: 6028.
DE
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
KW
OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-00200610.
PF
XX
XX 26-FEB-1999; 99US-0122487P.
PR
XX (GBST) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX WPI; 2000-500381/45.
XX
XX N-PSDE; AAC01953.
DR
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 13; SEQ ID NO 6028; 71pp + Sequence Listing; English.
PS
XX The present sequence is a polypeptide encoded by one of a large number of
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC tissues. EST sequences usually correspond mainly to the 3' untranslated
CC region (UTR) of the mRNA because they are often obtained from oligo-dT
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC sequences derived from the 5' ends of mRNAs and even in those cases where
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors
XX
XX Sequence 121 AA;
SQ
Query Match 61.3%; Score 46; DB 3; Length 121;
Best Local Similarity 54.5%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 PKPSPWAPKKH 11
Db 71 PDPTPEWPPQH 81
RESULT 15
ABU05501
ID ABU05501 standard; protein; 230 AA.
XX
XX AC ABU05501;
XX
XX 08-APR-2003 (first entry)
DT
XX
XX M. tuberculosis and M. leprae marker protein #152.
DE
XX Mycobacterioses; survival; virulence; protective antigen; vaccine;
KW mycobacterial disease; tuberculosis; leprosy.
KW
XX Mycobacterium tuberculosis.
OS
OS Mycobacterium leprae.
XX
XX WO2002/4903-A2.
XX
XX 26-SEP-2002.
PD

XX 22-FEB-2002; 2002WO-IB001973.
PF XX
XX 22-FEB-2001; 2001US-0270123P.
PR XX
XX (INSP) INST PASTEUR.
PA XX
XX Cole S;
PI XX
XX WPI; 2002-759885/82.
DR XX
XX Identifying and selecting genes for survival or virulence of mycobacteria
PT by a comparative genomic analysis of the sequences of Mycobacterium
PT tuberculosis and M. leprae.
XX
PS Claim 17; Page 322-323; 874pp; English.
XX
XX This invention relates to a novel method for identifying essential genes
CC for survival or virulence of mycobacteria species. The method comprises
CC aligning the genomic sequence of a first mycobacterium species on a
CC genomic sequence of a second mycobacterium species and selecting a
CC polynucleotide sequence that is highly conserved in both genomes with no
CC counterparts in other bacterial genomic sequences and that corresponds to
CC an essential gene for the survival or virulence of mycobacterium species.
CC The method of the invention is useful for detecting M. tuberculosis or M.
CC leprae infection. The method reduces the number of potential new targets
CC and protective antigens for new drugs and vaccine compositions to treat
CC and prevent mycobacterial diseases, particularly tuberculosis and
CC leprosy. The present sequence represents a marker protein from
CC Mycobacterium tuberculosis and Mycobacterium leprae identified using the
CC method of the invention
XX
SQ Sequence 230 AA;

Query Match 61.3%; Score 46; DB 5; Length 230;
Best Local Similarity 77.8%; Pred. NO. 81;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 PKPSPWAPK 9
Db 110 PKPSKWAPR 118
|||||

Search completed: May 27, 2004, 16:18:30
Job time : 34.2308 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:15:59 ; Search time 9.69231 Seconds
(without alignments)
63.918 Million cell updates/sec

Title: US-10-008-524A-168
Perfect score: 75
Sequence: 1 PKPSPWAPKKHR 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	100.0	12	4	US-09-314-268-168
2	75	100.0	15	4	US-09-314-268-167
3	75	100.0	20	2	US-08-934-915-40
4	75	100.0	20	2	US-08-934-915-170
5	75	100.0	25	4	US-09-314-268-4
6	75	100.0	30	2	US-08-934-915-42
7	75	100.0	92	4	US-09-314-268-90
8	68	90.7	35	4	US-09-314-268-117
9	52	69.3	8	4	US-09-314-268-37
10	52	69.3	8	4	US-09-314-268-36
11	51	68.0	8	4	US-09-314-268-118
12	51	68.0	34	4	US-09-314-268-118
13	50	66.7	8	4	US-09-314-268-34
14	50	66.7	8	4	US-09-314-268-35
15	49	65.3	325	4	US-09-921-099A-21
16	48	64.0	143	4	US-09-252-991A-21367
17	45	60.0	8	4	US-09-314-268-32
18	45	60.0	8	4	US-09-314-268-38
19	45	60.0	39	4	US-09-314-268-119
20	44	58.7	97	4	US-09-252-991A-20580
21	44	58.7	299	4	US-09-252-991A-17588
22	43	57.3	32	4	US-09-205-258-910
23	43	57.3	524	2	US-08-928-692-12
24	43	57.3	524	4	US-09-339-972-12
25	43	57.3	589	4	US-09-665-479A-4
26	43	57.3	959	4	US-09-252-991A-23758
27	42	56.0	40	4	US-09-314-268-106

ALIGNMENTS

RESULT 1

US-09-314-268-168
; Sequence 168, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:

; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; EARLIER FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 168
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-09-314-268-168

Query Match 100.0%; Score 75; DB 4; Length 12;

Best Local Similarity 100.0%; Pred. No. 7.5e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12

Db 1 PKPSPWAPKKHR 12

RESULT 2

US-09-314-268-167
; Sequence 167, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:

; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; EARLIER FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 167
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-09-314-268-167

Query Match 100.0%; Score 75; DB 4; Length 15;
 Best Local Similarity 100.0%; Pred. No. 9.4e-05;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 4 PKPSPWAPKKHR 15

RESULT 3
 US-08-934-915-40
 ; Sequence 40, Application US/08934915
 ; Patent No. 5932412
 ; GENERAL INFORMATION:
 ; APPLICANT: DILLNER, JOAKIM
 ; APPLICANT: DILLNER, LENA
 ; APPLICANT: CHENG, HWEE-MING
 ; TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
 ; TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,
 ; TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
 ; TITLE OF INVENTION: USEFUL IN IMMUNOASSAY FOR
 ; TITLE OF INVENTION: DIAGNOSTIC PURPOSES
 ; NUMBER OF SEQUENCES: 193
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MASON & ASSOCIATES, P.A.
 ; STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
 ; CITY: CLEARWATER
 ; STATE: FLORIDA
 ; COUNTRY: U.S.A.

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: Windows 3.0
 SOFTWARE: Microsoft Word 6.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/934,915
 FILING DATE: 22-SEP-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/949,836
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: LOUISE A. Foutch
 REGISTRATION NUMBER: 37,133
 REFERENCE/DOCKET NUMBER: 1946.6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 813-538-3800
 TELEFAX: 813-538-3820
 TELEX:

INFORMATION FOR SEQ ID NO: 40:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 20 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-934-915-40

Query Match 100.0%; Score 75; DB 2; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.00012;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 2 PKPSPWAPKKHR 13

RESULT 4
 US-08-934-915-170
 ; Sequence 170, Application US/08934915
 ; Patent No. 5932412
 ; GENERAL INFORMATION:
 ; APPLICANT: DILLNER, JOAKIM

APPLICANT: DILLNER, LENA
 APPLICANT: CHENG, HWEE-MING
 TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
 TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,
 TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
 TITLE OF INVENTION: USEFUL IN IMMUNOASSAY FOR
 TITLE OF INVENTION: DIAGNOSTIC PURPOSES
 NUMBER OF SEQUENCES: 193
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MASON & ASSOCIATES, P.A.
 STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
 CITY: CLEARWATER
 STATE: FLORIDA
 COUNTRY: U.S.A.

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: Windows 3.0
 SOFTWARE: Microsoft Word 6.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/934,915
 FILING DATE: 22-SEP-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/949,836
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: LOUISE A. Foutch
 REGISTRATION NUMBER: 37,133
 REFERENCE/DOCKET NUMBER: 1946.6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 813-538-3800
 TELEFAX: 813-538-3820
 TELEX:

INFORMATION FOR SEQ ID NO: 170:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 20 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-934-915-170

Query Match 100.0%; Score 75; DB 2; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.00012;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 2 PKPSPWAPKKHR 13

RESULT 5
 US-09-314-268-4
 ; Sequence 4, Application US/09314268
 ; Patent No. 6346377
 ; GENERAL INFORMATION:
 ; APPLICANT: Doobar, John
 ; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
 ; TITLE OF INVENTION: VIRUSES
 ; FILE REFERENCE: 3789/80902
 ; CURRENT APPLICATION NUMBER: US/09/314,268
 ; CURRENT FILING DATE: 1999-03-19
 ; EARLIER APPLICATION NUMBER: 09/314,268
 ; EARLIER FILING DATE: 1999-05-18
 ; NUMBER OF SEQ ID NOS: 179
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 25
 ; TYPE: PRT
 ; ORGANISM: Human papillomavirus type 16
 US-09-314-268-4

Query Match 100.0%; Score 75; DB 4; Length 25;

Best Local Similarity 100.0%; Pred. No. 0.00016; Gaps 0;
Matches 12; Conservative 0; Mismatches 0; Indels 0;

Qy 1 KPSPWAPKKHR 12
Db 4 KPSPWAPKKHR 15

RESULT 6

US-08-934-915-42
; Sequence 42, Application US/08934915
; Patent No. 5932412
; GENERAL INFORMATION:
; APPLICANT: DILLNER, JOAKIM
; APPLICANT: DILLNER, LENA
; APPLICANT: CHENG, HWEE-MING
; TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
; TITLE OF INVENTION: PAPILLOMAVIRUS 1, 5, 6, 8,
; TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
; TITLE OF INVENTION: USEFUL IN IMMUNOASSAY FOR
; TITLE OF INVENTION: DIAGNOSTIC PURPOSES
; NUMBER OF SEQUENCES: 193
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MASON & ASSOCIATES, P.A.
; STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
; CITY: CLEARWATER
; STATE: FLORIDA
; COUNTRY: U.S.A.

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows 3.0
; SOFTWARE: Microsoft Word 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/934,915
; FILING DATE: 22-SEP-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/949,836
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: LOUISE A. Foutch
; REGISTRATION NUMBER: 37,133
; REFERENCE/DOCKET NUMBER: 1946.6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 813-538-3800
; TELEFAX: 813-538-3820
; TELEX:

; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-934-915-42

Query Match 100.0%; Score 75; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPWAPKKHR 12
Db 10 KPSPWAPKKHR 21

RESULT 7

US-09-314-268-90
; Sequence 90, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES

; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 90
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-09-314-268-90

Query Match 100.0%; Score 75; DB 4; Length 92;
Best Local Similarity 100.0%; Pred. No. 0.00057;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPWAPKKHR 12
Db 29 KPSPWAPKKHR 40

RESULT 8

US-09-314-268-117
; Sequence 117, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 117
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-09-314-268-117

Query Match 90.7%; Score 68; DB 4; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0022;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KPSPWAPKKHR 12
Db 1 KPSPWAPKKHR 11

RESULT 9

US-09-314-268-33
; Sequence 33, Application US/09314268
; Patent No. 6346377
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 33
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: octapeptide antigen
US-09-314-268-33

Query Match 69.3%; Score 52; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KPSPWAP 8
Db 1 KPSPWAP 8

RESULT 10

US-09-314-268-37
; Sequence 37, Application US/09314268

; Patent No. 6346377

; GENERAL INFORMATION:

; APPLICANT: Doorbar, John

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

; FILE REFERENCE: 3789/80902

; CURRENT APPLICATION NUMBER: US/09/314,268

; CURRENT FILING DATE: 1999-03-19

; EARLIER APPLICATION NUMBER: 09/314,268

; EARLIER FILING DATE: 1999-05-18

; NUMBER OF SEQ ID NOS: 179

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 37

; LENGTH: 8

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: octapeptide antigen

US-09-314-268-37

Query Match 69.3%; Score 52; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 PWAPKKHR 12
Db 1 PWAPKKHR 8

RESULT 11

US-09-314-268-36

; Sequence 36, Application US/09314268

; Patent No. 6346377

; GENERAL INFORMATION:

; APPLICANT: Doorbar, John

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

; FILE REFERENCE: 3789/80902

; CURRENT APPLICATION NUMBER: US/09/314,268

; CURRENT FILING DATE: 1999-03-19

; EARLIER APPLICATION NUMBER: 09/314,268

; EARLIER FILING DATE: 1999-05-18

; NUMBER OF SEQ ID NOS: 179

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 36

; LENGTH: 8

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: octapeptide antigen

US-09-314-268-36

Query Match 68.0%; Score 51; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SPWAPKKH 11
Db 1 SPWAPKKH 8

RESULT 12

US-09-314-268-118

; Sequence 118, Application US/09314268

; Patent No. 6346377

; GENERAL INFORMATION:

; APPLICANT: Doorbar, John

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

; FILE REFERENCE: 3789/80902

; CURRENT APPLICATION NUMBER: US/09/314,268

; CURRENT FILING DATE: 1999-03-19

; EARLIER APPLICATION NUMBER: 09/314,268

; EARLIER FILING DATE: 1999-05-18

; NUMBER OF SEQ ID NOS: 179

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 118

; LENGTH: 34

; TYPE: PRT

; ORGANISM: Human papillomavirus type 35

US-09-314-268-118

Query Match 68.0%; Score 51; DB 4; Length 34;
Best Local Similarity 72.7%; Pred. No. 0.55;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KPSPWAPKKHR 12
Db 1 KPAPWAPQKPR 11

RESULT 13

US-09-314-268-34

; Sequence 34, Application US/09314268

; Patent No. 6346377

; GENERAL INFORMATION:

; APPLICANT: Doorbar, John

; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

; FILE REFERENCE: 3789/80902

; CURRENT APPLICATION NUMBER: US/09/314,268

; CURRENT FILING DATE: 1999-03-19

; EARLIER APPLICATION NUMBER: 09/314,268

; EARLIER FILING DATE: 1999-05-18

; NUMBER OF SEQ ID NOS: 179

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 34

; LENGTH: 8

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: octapeptide antigen

US-09-314-268-34

Query Match 66.7%; Score 50; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 KPSPWAPK 9
Db 1 KPSPWAPK 8

RESULT 14

US-09-314-268-35

; Sequence 35, Application US/09314268

; Patent No. 6346377

; GENERAL INFORMATION:

; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE OF INVENTION: VIRUSES
; FILE REFERENCE: 3789/80902
; CURRENT APPLICATION NUMBER: US/09/314,268
; CURRENT FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/314,268
; EARLIER FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-09-314-268-35

Query Match 66.7%; Score 50; DB 4; Length 8;
Best Local Similarity 100.0%; Pred.No. 3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PSPWAPKK 10
| | | | | | | |
Db 1 PSPWAPKK 8

RESULT 15
US-09-921-099A-21
; Sequence 21, Application US/09921099A
; Patent No. 6602707
; GENERAL INFORMATION:
; APPLICANT: Hefeneider, Steven
; APPLICANT: Merkins, Louise
; APPLICANT: Bennett, Robert
; APPLICANT: Seiss, Donald
; TITLE OF INVENTION: Mammalian DNA Binding Membrane-Associated Protein-encoding Gene a
; FILE REFERENCE: 00-617-A
; CURRENT APPLICATION NUMBER: US/09/921,099A
; CURRENT FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-921-099A-21

Query Match 65.3%; Score 49; DB 4; Length 325;
Best Local Similarity 70.0%; Pred.No. 10;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKK 10
| : | | | | | |
Db 79 PRPSPWLPSK 88

Search completed: May 27, 2004, 16:22:37
Job time : 9.69231 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:21:04 ; Search time 25.3846 Seconds
(without alignments)
132.009 Million cell updates/sec

Title: US-10-008-524A-168
Perfect score: 75
Sequence: 1 PKPSPWAPKXHR 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1151071 seqs, 279249464 residues

Total number of hits satisfying chosen parameters: 1151071

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	100.0	12	14	US-10-008-524A-168
2	75	100.0	12	15	US-10-350-719-168
3	75	100.0	15	14	US-10-008-524A-167
4	75	100.0	15	15	US-10-350-719-167
5	75	100.0	25	14	US-10-008-524A-4
6	75	100.0	26	15	US-10-350-719-4
7	75	100.0	92	14	US-10-008-524A-90
8	75	100.0	92	15	US-10-350-719-90
9	68	90.7	35	15	US-10-008-524A-117
10	68	90.7	35	15	US-10-350-719-117
11	52	69.3	8	14	US-10-008-524A-33
12	52	69.3	8	14	US-10-008-524A-37
13	52	69.3	8	15	US-10-350-719-33
14	52	69.3	8	15	US-10-350-719-37
15	51	68.0	8	14	US-10-008-524A-36

16	51	68.0	8	15	US-10-350-719-36	Sequence 36, Appl
17	51	68.0	34	14	US-10-008-524A-118	Sequence 118, App
18	51	68.0	34	15	US-10-350-719-118	Sequence 118, App
19	50	66.7	8	14	US-10-008-524A-34	Sequence 34, Appl
20	50	66.7	8	14	US-10-008-524A-35	Sequence 35, Appl
21	50	66.7	8	15	US-10-350-719-34	Sequence 34, Appl
22	50	66.7	8	15	US-10-350-719-35	Sequence 35, Appl
23	50	66.7	113	12	US-10-424-599-254084	Sequence 254084,
24	46	61.3	93	12	US-10-424-599-264398	Sequence 264398,
25	46	61.3	230	14	US-10-080-170-152	Sequence 152, App
26	45	60.0	8	14	US-10-008-524A-32	Sequence 32, Appl
27	45	60.0	8	14	US-10-008-524A-38	Sequence 38, Appl
28	45	60.0	8	15	US-10-350-719-32	Sequence 32, Appl
29	45	60.0	8	15	US-10-350-719-38	Sequence 38, Appl
30	45	60.0	39	14	US-10-008-524A-119	Sequence 119, App
31	45	60.0	39	15	US-10-350-719-119	Sequence 119, App
32	45	60.0	132	12	US-10-425-114-71639	Sequence 71639, A
33	45	60.0	234	12	US-10-424-599-266133	Sequence 266133,
34	44	58.7	56	12	US-10-424-599-233479	Sequence 233479,
35	44	58.7	114	14	US-10-029-386-30822	Sequence 30822, A
36	44	58.7	142	15	US-10-131-487A-125	Sequence 125, App
37	44	58.7	168	15	US-10-131-487A-192	Sequence 192, App
38	44	58.7	174	16	US-10-389-566-2059	Sequence 2059, App
39	44	58.7	711	15	US-10-369-493-3168	Sequence 3168, App
40	44	58.7	783	15	US-10-150-559-6	Sequence 6, Appli
41	44	58.7	783	15	US-10-440-352-6	Sequence 6, Appli
42	43	57.3	32	10	US-09-933-767-910	Sequence 910, App
43	43	57.3	32	12	US-10-004-860-910	Sequence 910, App
44	43	57.3	32	14	US-10-023-282-910	Sequence 910, App
45	43	57.3	56	10	US-09-764-891-3041	Sequence 3041, App

ALIGNMENTS

RESULT 1

US-10-008-524A-168
; Sequence 168, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 168
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-168

Query Match 100.0%; Score 75; DB 14; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSPWAPKXHR 12
|||||
Db 1 PKPSPWAPKXHR 12

RESULT 2

US-10-350-719-168
; Sequence 168, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES

```
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GH01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 168
; LENGTH: 12
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-350-719-168

Query Match      100.0%; Score 75; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 PKPSPWAPKKHR 12
    |||||
Db  1 PKPSPWAPKKHR 12
    |||||

RESULT 3
US-10-008-524A-167
; Sequence 167, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 167
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-167

Query Match      100.0%; Score 75; DB 14; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 PKPSPWAPKKHR 12
    |||||
Db  4 PKPSPWAPKKHR 15
    |||||

RESULT 4
US-10-350-719-167
; Sequence 167, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GH01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 167
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-167

Query Match      100.0%; Score 75; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 PKPSPWAPKKHR 12
    |||||
Db  4 PKPSPWAPKKHR 15
    |||||

RESULT 5
US-10-008-524A-4
; Sequence 4, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-4

Query Match      100.0%; Score 75; DB 14; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.0036;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 PKPSPWAPKKHR 12
    |||||
Db  4 PKPSPWAPKKHR 15
    |||||

RESULT 6
US-10-350-719-4
; Sequence 4, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GH01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Human papillomavirus type 16
US-10-350-719-4

Query Match      100.0%; Score 75; DB 15; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.0037;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  1 PKPSPWAPKKHR 12
    |||||
Db  4 PKPSPWAPKKHR 15
    |||||
```

```

RESULT 7
US-10-008-524A-90
; Sequence 90, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; TYPE: PRT
; LENGTH: 92
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-90

Query Match          100.0%; Score 75; DB 14; Length 92;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPWAPKKHR 12
Db 29 KPSPWAPKKHR 40

RESULT 8
US-10-350-719-90
; Sequence 90, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; TYPE: PRT
; LENGTH: 92
; ORGANISM: Human papillomavirus type 16
US-10-350-719-90

Query Match          100.0%; Score 75; DB 15; Length 92;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KPSPWAPKKHR 12
Db 29 KPSPWAPKKHR 40

RESULT 9
US-10-008-524A-117
; Sequence 117, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; TYPE: PRT
; LENGTH: 8
; ORGANISM: Artificial Sequence
; FEATURE:

```

```

; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 117
; TYPE: PRT
; LENGTH: 35
; ORGANISM: Human papillomavirus type 16
US-10-008-524A-117

Query Match          90.7%; Score 68; DB 14; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KPSPWAPKKHR 12
Db 1 KPSPWAPKKHR 11

RESULT 10
US-10-350-719-117
; Sequence 117, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 117
; TYPE: PRT
; LENGTH: 35
; ORGANISM: Human papillomavirus type 16
US-10-350-719-117

Query Match          90.7%; Score 68; DB 15; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KPSPWAPKKHR 12
Db 1 KPSPWAPKKHR 11

RESULT 11
US-10-008-524A-33
; Sequence 33, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; TYPE: PRT
; LENGTH: 8
; ORGANISM: Artificial Sequence
; FEATURE:

```

```

; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-10-008-524A-33

Query Match      69.3%; Score 52; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 PKPSPWAP 8
      |||||
Db      1 PKPSPWAP 8

RESULT 12
US-10-008-524A-37
; Sequence 37, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-10-008-524A-37

Query Match      69.3%; Score 52; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 PWAPKKHR 12
      |||||
Db      1 PWAPKKHR 8

RESULT 13
US-10-350-719-33
; Sequence 33, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-10-350-719-33

Query Match      69.3%; Score 52; DB 15; Length 8;

```

```

Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 PKPSPWAP 8
      |||||
Db      1 PKPSPWAP 8

RESULT 14
US-10-350-719-37
; Sequence 37, Application US/10350719
; Publication No. US20030219726A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/2162
; CURRENT APPLICATION NUMBER: US/10/350,719
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: PCT/GB01/01176
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: GB0018140.4
; PRIOR FILING DATE: 2000-07-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-10-350-719-37

Query Match      69.3%; Score 52; DB 15; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      5 PWAPKKHR 12
      |||||
Db      1 PWAPKKHR 8

RESULT 15
US-10-008-524A-36
; Sequence 36, Application US/10008524A
; Publication No. US20030175682A1
; GENERAL INFORMATION:
; APPLICANT: Doorbar, John
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA
; TITLE OF INVENTION: VIRUSES
; FILE REFERENCE: 18396/1074
; CURRENT APPLICATION NUMBER: US/10/008,524A
; CURRENT FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: 09/314,268
; PRIOR FILING DATE: 1999-05-18
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: octapeptide antigen
US-10-008-524A-36

Query Match      68.0%; Score 51; DB 14; Length 8;
Best Local Similarity 100.0%; Pred. No. 1e+06;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 SPWAPKKH 11
      |||||

```


Fri May 28 09:17:11 2004

us-10-008-524a-168.rapb

Page 5

Db 1 SPWAPKX 8

Search completed: May 27, 2004, 16:34:19
Job time : 26.3846 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:13:48 ; Search time 8.07692 Seconds
(without alignments)
142.913 Million cell updates/sec

Title: US-10-008-524A-168

Perfect score: 75

Sequence: 1 PKPSPWAPKKHR 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.*

1: P1r1.*

2: P1r2.*

3: P1r3.*

4: P1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	100.0	95	1 W4WLHS	E4 protein - human
2	58	77.3	96	1 W4WL35	E4 protein - human
3	52	69.3	102	1 W4WL31	E4 protein - human
4	51	68.0	1060	2 A10201	beta-galactosidase
5	49	65.3	325	2 S57977	CCCH zinc finger p
6	48	64.0	992	2 T38817	hypothetical prote
7	47	62.7	378	1 A40004	histidine decarbox
8	46	61.3	230	2 S72714	Lepb170 F2 64 pro
9	45	60.0	87	1 W4WL51	E4 protein - human
10	45	60.0	1553	2 T03301	rab3 effector prot
11	44	58.7	88	1 W4WL18	E4 protein - human
12	43	57.3	393	2 C64613	conserved hypochet
13	43	57.3	400	2 D71900	hypothetical prote
14	43	57.3	524	2 A31318	glucose transporte
15	43	57.3	721	2 C84677	probable membrane
16	43	57.3	910	2 JC4609	chitin synthase (E
17	43	57.3	925	2 T07713	probable ABC-type
18	42	56.0	143	2 D83148	hypothetical prote
19	42	56.0	197	2 F82029	probable periplasm
20	42	56.0	210	2 B81008	hypothetical prote
21	42	56.0	330	2 T05717	probable extensin
22	42	56.0	389	2 T05782	hypothetical prote
23	42	56.0	474	2 T34193	G protein-coupled
24	42	56.0	687	2 A57713	chloride channel C
25	42	56.0	984	2 JC5545	integrin beta-4 pr
26	42	56.0	1875	2 A36429	integrin beta-4 ch
27	41	54.7	232	2 AC2621	conserved hypochet
28	41	54.7	232	2 C97403	hemK protein homol
29	41	54.7	243	2 S19148	thiogalactosidase (E

ALIGNMENTS

RESULT 1

W4WLHS

E4 protein - human papillomavirus type 16

C/Species: human papillomavirus type 16

C/Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 20-Aug-1999

C/Accession: A22355; T10425

R/Seedorf, K.; Krammer, G.; Durst, M.; Suhai, S.; Rowekamp, W.G.

Virology 145, 181-185, 1985

A/Title: Human papillomavirus type 16 DNA sequence.

A/Reference number: A22355; MUID:85246220; PMID:2990099

A/Accession: A22355

A/Molecule type: DNA

A/Residues: 1-95 <SER>

A/Cross-references: GB:K02718; NID:g333031; PIDN:AAA46937.1; PID:g459913

R/Kennedy, I.M.; Haddow, J.K.; Clements, J.B.

J. Virol. 65, 2093-2097, 1991

A/Title: A negative element in the human poapillomavirus type 16 genome acts at the lev

A/Reference number: Z17014; MUID:91162763; PMID:1848319

A/Accession: T10425

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-95 <KEN>

A/Cross-references: EMBL:K02718; NID:g333031; PIDN:AAA46937.1; PID:g459913

C/Genetics:

A/Gene: E4

C/Superfamily: papillomavirus E4 protein

C/Keywords: early protein

Query Match 100.0%; Score 75; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 0.00035;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PKPSPWAPKKHR 12

Db 32 PKPSPWAPKKHR 43

RESULT 2

W4WL35

E4 protein - human papillomavirus type 35

C/Species: human papillomavirus type 35

A/Note: host Homo sapiens (man)

C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 27-Jan-1995

C/Accession: C40824

R/Marich, J.E.; Pontesler, A.V.; Rice, S.M.; McGraw, K.A.; Dubensky, T.W.

Virology 186, 770-776, 1992

A/Title: The phylogenetic relationship and complete nucleotide sequence of human papill

A/Reference number: A40824; MUID:92124753; PMID:1310198

A/Accession: C40824

A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 1-96 <MAR>

A;Cross-references: GB:M74117
C;Superfamily: papillomavirus E4 protein
C;Keywords: early protein

Query Match 77.3%; Score 58; DB 1; Length 96;

Best Local Similarity 75.0%; Pred. No. 0.1;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKH 12

Db 32 PKPAPWAPQKPR 43

RESULT 3

W4WL31
E4 protein - human papillomavirus type 31

C;Species: human papillomavirus type 31

A;Note: host Homo sapiens (man)

C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999

C;Accession: E32444

R;Goldsbrough, M.D.; DiSilvestre, D.; Temple, G.F.; Lorincz, A.T.

Virolology 171, 306-311, 1989

A;Title: Nucleotide sequence of human papillomavirus type 31: a cervical neoplasia-associated

A;Reference number: A94398; MUID:89299478; PMID:2545036

A;Accession: E32444

A;Status: translation not shown

A;Molecule type: DNA

A;Residues: 1-102 <GOL>

A;Cross-references: GB:J04353; NID:9333048; PIDN:AAA46949.1; PID:9459915

A;Note: in GenBank entry PPH31A the initiation codon UUG for residue 1 is translated as

C;Superfamily: papillomavirus E4 protein

C;Keywords: early protein

Query Match 69.3%; Score 52; DB 1; Length 102;

Best Local Similarity 80.0%; Pred. No. 0.81;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKK 10

Db 33 PKPAPWAPVK 42

RESULT 4

AI0201

beta-galactosidase (EC 3.2.1.23) [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001

C;Accession: AI0201

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AI0201

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1060 <KUR>

A;Cross-references: GB:AL590842; PIDN:CAC90476.1; PID:gl5979691; GSPDB:GN00175

C;Genetics:

A;Gene: lacZ

C;Superfamily: beta-galactosidase

C;Keywords: glycosidase; hydrolase

Query Match 68.0%; Score 51; DB 2; Length 1060;

Best Local Similarity 58.3%; Pred. No. 9.9;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKH 12

Db 723 PKETWSPAQRH 734

RESULT 7

A40004

histidine decarboxylase (EC 4.1.1.22) - Enterobacter aerogenes

C;Species: Enterobacter aerogenes

RESULT 5

S57977

CCCH zinc finger protein CTH1 - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein YD8358.07c; protein YDR151c

C;Species: Saccharomyces cerevisiae

C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999

C;Accession: S57977; JC5001

R;Murphy, L.; Richards, C.; Harris, D.

submitted to the EMBL Data Library, July 1995

A;Reference number: S57971

A;Accession: S57977

A;Molecule type: DNA

A;Residues: 1-325 <MUR>

A;Cross-references: EMBL:Z50046; NID:g899393; PIDN:CAA90373.1; PID:g899400; MIPS:YDR151

A;Experimental source: strain AB972

R;Thompson, M.J.; Lai, W.S.; Taylor, G.A.; Blackshear, P.J.

Gene 174, 225-233, 1996

A;Title: Cloning and characterization of two yeast genes encoding members of the CCCH

A;Reference number: JC5001; MUID:97045817; PMID:8890739

A;Accession: JC5001

A;Molecule type: DNA

A;Residues: 1-141,'RV',144-325 <THO>

A;Cross-references: GB:I42133; NID:gl020082; PIDN:AAB39897.1; PID:gl020083

C;Comment: This protein belongs to the CCCH-type zinc finger protein family, and is a

C;Genetics:

A;Gene: SGD:CTH1

A;Cross-references: SGD:S0002558; MIPS:YDR151c

A;Map position: 4R

C;Keywords: zinc finger

F;132-136/Region: nuclear location signal

Query Match 65.3%; Score 49; DB 2; Length 325;

Best Local Similarity 70.0%; Pred. No. 6.5;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKK 10

Db 79 PRPSFWLPK 88

RESULT 6

T38817

hypothetical protein SPAC4F10.13c - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C;Accession: T38817

R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Raebandream, M.A.; Wood, V.

submitted to the EMBL Data Library, September 1997

A;Reference number: Z21813

A;Accession: T38817

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-992 <CON>

A;Cross-references: EMBL:Z99890; NID:el060691; PIDN:CABL1716.1; GSPDB:GN00066; SPDB:SPA

A;Experimental source: strain 972h-; cosmid c4F10

C;Genetics:

A;Gene: SPDB:SPAC4F10.13c

A;Map position: 1

A;Introns: 13/2

Query Match 64.0%; Score 48; DB 2; Length 992;

Best Local Similarity 69.2%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 1 PKPSPW--APKKH 11

Db 622 PKPSPWKSLLPKH 634

C;Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 18-Jun-1999
 C;Accession: A40004
 R;Kamath, A.V.; Vaaler, G.L.; Snell, E.E.
 J. Biol. Chem. 266, 9432-9437, 1991
 A;Title: Pyridoxal phosphate-dependent histidine decarboxylases. Cloning, sequencing, and enzymes.
 A;Reference number: A40004; MUID:91236707; PMID:2033044
 A;Accession: A40004
 A;Status: not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-378 <KAM>
 A;Cross-references: GB:M62745; NID:g435593; PIDN:AAA24802.1; PID:g435594
 C;Superfamily: Klebsiella histidine decarboxylase
 C;Keywords: carbon-carbon lyase; carboxy-lyase; phosphoprotein; pyridoxal phosphate
 P;233/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 62.7%; Score 47; DB 1; Length 378;
 Best Local Similarity 72.7%; Pred. No. 15;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PKPSPWAPKKH 11
 |||||
 Db 329 PKPSEWVWKH 339

RESULT 8
 S72714
 Lep1170_F2_64 protein - Mycobacterium leprae
 C;Species: Mycobacterium leprae
 C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 C;Accession: S72714
 R;Smith, D.R.; Robinson, K.
 submitted to the EMBL Data Library, November 1993
 A;Description: Mycobacterium leprae cosmid B1170.
 A;Reference number: S72693
 A;Accession: S72714
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-230 <SMI>
 A;Cross-references: EMBL:U00010; NID:g466780; PIDN:AAAL7078.1; PID:g466802
 C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1517

Query Match 61.3%; Score 46; DB 2; Length 230;
 Best Local Similarity 77.8%; Pred. No. 13;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSPWAPK 9
 |||||
 Db 110 PKPSKWAPR 118

RESULT 9
 W4WL51
 E4 protein - human papillomavirus type 51
 C;Species: human papillomavirus type 51
 A;Note: host Homo sapiens (man)
 C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 27-Jan-1995
 C;Accession: C40415
 R;Lungu, O.; Crum, C.P.; Silverstein, S.J.
 J. Virol. 65, 4216-4225, 1991
 A;Title: Biologic properties and nucleotide sequence analysis of human papillomavirus type 51.
 A;Reference number: A40415; MUID:91303675; PMID:1649326
 A;Accession: C40415
 A;Status: translation not shown
 A;Molecule type: DNA
 A;Residues: 1-87 <LUN>
 A;Cross-references: GB:M62877
 C;Superfamily: papillomavirus E4 protein
 C;Keywords: early protein

Query Match 60.0%; Score 45; DB 1; Length 87;
 Best Local Similarity 66.7%; Pred. No. 7.3;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PKPSPWAPKKH 12
 |||||
 Db 27 FLPPAWAPKKR 38

RESULT 10
 T03301
 rab3 effector protein Rim - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
 C;Accession: T03301
 R;Wang, Y.; Okamoto, M.; Schmitz, F.; Hofmann, K.; Sudhof, T.C.
 Nature 388, 593-598, 1997
 A;Title: Rim is a putative Rab3 effector in regulating synaptic-vesicle fusion.
 A;Reference number: Z14897; MUID:97394473; PMID:92552191
 A;Accession: T03301
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1553 <WAN>
 A;Cross-references: EMBL:AF007836; NID:g2317777; PIDN:AAB66703.1; PID:g2317778
 A;Experimental source: tissue-type brain
 C;Genetics:
 A;Note: RIM
 C;Function:
 A;Description: Rim protein is proposed as Rab3-dependent regulator of synaptic-vesicle C;Keywords: GTP binding; zinc finger

Query Match 60.0%; Score 45; DB 2; Length 1553;
 Best Local Similarity 54.5%; Pred. No. 1.1e+02;
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSPWAPKKH 11
 |||||
 Db 892 PQSPFMPERRH 902

RESULT 11
 W4WL18
 E4 protein - human papillomavirus type 18
 C;Species: human papillomavirus type 18
 C;Date: 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 16-Jul-1999
 C;Accession: E26251
 R;Cole, S.T.; Danos, O.
 J. Mol. Biol. 193, 539-608, 1987
 A;Title: Nucleotide sequence and comparative analysis of the human papillomavirus type 18.
 A;Reference number: A92937; MUID:87283882; PMID:3039146
 A;Accession: E26251
 A;Molecule type: DNA
 A;Residues: 1-88 <COL>
 A;Cross-references: GB:X05015; NID:g60975; PIDN:CAA28668.1; PID:g60980
 C;Superfamily: papillomavirus E4 protein
 C;Keywords: early protein

Query Match 58.7%; Score 44; DB 1; Length 88;
 Best Local Similarity 60.0%; Pred. No. 10;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSPWAPKK 10
 |||||
 Db 29 PAPCPWAPQR 38

RESULT 12
 C64613
 conserved hypothetical protein HP0747 - Helicobacter pylori (strain 26695)
 C;Species: Helicobacter pylori
 C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C;Accession: C64613
 R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.
A;Reference number: A64320; MUID:97394467; PMID:9252185
A;Accession: C64613
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-393 <TOM>
A;Cross-references: GB:AE000511; NID:g2313869; PIDN:AAD07796.1; PID:g231387

Query Match 57.3%; Score 43; DB 2; Length 393;
Best Local Similarity 70.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PSPWAPKKHR 12
| | | | |
Db 197 PVPWNEKKHR 206

RESULT 13
D71900
hypothetical protein jhp0684 - *Helicobacter pylori* (strain J99)
C;Species: *Helicobacter pylori*
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C;Accession: D71900
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: D71900
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-400 <ARN>
A;Cross-references: GB:AE001500; GB:AE001439; NID:g4155238; PIDN:AAD06272.1; PID:g415525
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp0684

Query Match 57.3%; Score 43; DB 2; Length 400;
Best Local Similarity 70.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PSPWAPKKHR 12
| | | | |
Db 197 PVPWNEKKHR 206

RESULT 14
A31318
glucose transporter-like protein - human
C;Species: *Homo sapiens* (man)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 24-Sep-1999
C;Accession: A31318
R;Fukumoto, H.; Seino, S.; Imura, H.; Seino, Y.; Eddy, R.L.; Fukushima, Y.; Byers, M.G.;
Proc. Natl. Acad. Sci. U.S.A. 85, 5434-5438, 1988
A;Title: Sequence, tissue distribution, and chromosomal localization of mRNA encoding a
A;Reference number: A31318; MUID:88289735; PMID:3399500
A;Accession: A31318
A;Molecule type: mRNA
A;Residues: 1-524 <PUK>
A;Cross-references: GB:J03810; NID:g187133; PIDN:AAA59514.1; PID:g307125
C;Genetics:
A;Gene: GDB:SLC2A2; GLUT2
A;Cross-references: GDB:119995; OMIM:138160
A;Map position: 3q26.2-3q27
C;Superfamily: glucose transport protein
C;Keywords: transmembrane protein

Query Match 57.3%; Score 43; DB 2; Length 524;
Best Local Similarity 60.0%; Pred. No. 75;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSPWAPKK 10
| | | | |
Db 76 PKPTWAESEE 85

RESULT 15

C84677
probable membrane transporter [imported] - *Arabidopsis thaliana*
C;Species: *Arabidopsis thaliana* (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: C84677
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: C84677
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-721 <STO>
A;Cross-references: GB:AE002093; NID:g3860251; PIDN:AACT3019.1; GSPDB:GNC00139
C;Genetics:
A;Gene: Atg27810
A;Map position: 2

Query Match 57.3%; Score 43; DB 2; Length 721;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSPWAP 8
| | | | |
Db 10 PKPGPWPP 17

Search completed: May 27, 2004, 16:21:44
Job time : 8.07692 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:03:58 ; Search time 5.53846 Seconds
(without alignments)
112.819 Million cell updates/sec

Title: US-10-008-524A-168
Perfect score: 75
Sequence: 1 PKPSPWAPKKHR 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	100.0	95	VE4_HPV16	P06922 human papil
2	58	77.3	96	VE4_HPV35	P27224 human papil
3	52	69.3	102	VE4_HPV31	P17384 human papil
4	49	65.3	325	CTH1_YEAST	P47976 saccharomyc
5	47	62.7	377	DCHS_ENTAE	P28577 enterobacte
6	45	60.0	87	VE4_HPV51	P26548 human papil
7	45	60.0	461	Y514_HUMAN	O60269 homo sapien
8	45	60.0	1615	RIM1_RAT	Q9J174 rattus norv
9	45	60.0	1692	RIM1_HUMAN	Q86ur5 homo sapien
10	44	58.7	88	VE4_HPV18	P06791 human papil
11	44	58.7	174	RL22_SECE	Q06036 secale cere
12	43	57.3	393	TRMB_HELPY	O25443 helicobacte
13	43	57.3	400	TRMB_HELPJ	Q92196 helicobacte
14	43	57.3	524	GTR2_HUMAN	P11168 homo sapien
15	43	57.3	910	CHSA_AMPQU	Q12564 ampelomyc
16	42	56.0	1822	ITB4_HUMAN	P26144 homo sapien
17	41	54.7	232	TRMB_AGR15	Q8u1d4 agrobacteri
18	41	54.7	243	MYR2_SINAL	Q98b73 rhizobium 1
19	41	54.7	247	TRMB_RHIL0	Q98b73 rhizobium 1
20	41	54.7	286	TRMB_BIFL0	O8g3t4 bifidobacte
21	41	54.7	435	FUT9_ARATH	Q8xi77 arabidopsis
22	41	54.7	471	YHJ6_YEAST	P38770 saccharomyc
23	41	54.7	670	ZN16_HUMAN	P17020 homo sapien
24	41	54.7	778	SYQ_DROME	Q9y105 drosophila
25	41	54.7	1236	POLS_WEEV	P13897 western equ
26	41	54.7	1268	VGLN_HUMAN	Q00341 homo sapien
27	41	54.7	1270	VGLN_CHICK	P81021 gallus gall
28	40	53.3	163	CU69_HUMAN	P58556 homo sapien
29	40	53.3	202	YF56_MYCTU	Q10774 mycobacteri
30	40	53.3	282	SP0J_BACSU	P26497 bacillus su
31	40	53.3	437	CAH9_MOUSE	Q8vbb5 mus musculu
32	40	53.3	478	BM3B_HUMAN	P55107 homo sapien
33	40	53.3	542	ZYX_CHICK	Q04584 gallus gall

```

34 40 53.3 564 1 PSL1_HUMAN      Q8tct7 homo sapien
35 40 53.3 806 1 MK07_MOUSE     Q9wvs8 mus musculu
36 40 53.3 815 1 MK07_HUMAN     Q13164 homo sapien
37 39.5 52.7 971 1 OPAL_ONCWA     O93248 oncorhynchu
38 39 52.0 205 1 YNP2_CABEL     P34555 caenorhabdi
39 39 52.0 251 1 CBX7_HUMAN      O95931 homo sapien
40 39 52.0 305 1 HIS7_NEIMA      Q9j1r7 neisseria m
41 39 52.0 345 1 FXL1_HUMAN      Q12952 homo sapien
42 39 52.0 495 1 Y892_MYCTU      Q10532 mycobacteri
43 39 52.0 533 1 PPO3_HUMAN      Q9v6f1 homo sapien
44 39 52.0 561 1 3BP2_HUMAN      P78314 homo sapien
45 39 52.0 561 1 SR72_SCHPO      OS9787 schizosacch

```

ALIGNMENTS

```

RESULT 1
VE4_HPV16
ID   VE4_HPV16          STANDARD;          PRT;          95 AA.
AC   P06922;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-JUL-1993 (Rel. 26, Last annotation update)
DE   Probable E4 protein.
OS   Human papillomavirus type 16.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10581;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85246220; PubMed=2990099;
RA   Seedorf K., Kramer G., Durst M., Suhai S., Rowekamp W.G.;
RT   "Human papillomavirus type 16 DNA sequence.";
RL   Virology 145:181-185(1985).
RN   [2]
RP   SEQUENCE OF 6-95 FROM N.A.
RX   MEDLINE=90218027; PubMed=2157796;
RA   Schneider-Maunoury S., Pehau-Arnauudet G., Breitburd F., Orth G.;
RT   "Expression of the human papillomavirus type 16 genome in SK-v cells,
RT   a line derived from a vulvar intraepithelial neoplasia.";
RL   J. Gen. Virol. 71:809-817(1990).
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; K02718; AAA46937.1; -.
DR   EMBL; D00735; BAA00634.1; -.
DR   PIR; A22355; W4WLHS.
DR   InterPro; IPR003861; Papilloma_E4.
DR   Pfam; PF02711; Pap_E4; 1.
KW   Early protein.
SQ   SEQUENCE 95 AA; 10594 MW; AED4269D177307CE CRC64;

```

```

Query Match      100.0%; Score 75; DB 1; Length 95;
Best Local Similarity 100.0%; Pred. No. 0.00024;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 PKPSPWAPKKHR 12
DB 32 PKPSPWAPKKHR 43

```

```

RESULT 2
VE4_HPV35
ID   VE4_HPV35          STANDARD;          PRT;          96 AA.
AC   P27224;
DT   01-AUG-1992 (Rel. 23, Created)

```

```

DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 35.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10587;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92124753; PubMed=1310198;
RA Marich J.E., Pontsler A.V., Rice S.M., McGraw K.A., Dubensky T.W.;
RT "The phylogenetic relationship and complete nucleotide sequence of
RL human papillomavirus type 35.";
RL Virology 186:770-776(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M74117; AAA46965.2; -.
DR PIR; C40824; W4WL35.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.
SQ SEQUENCE 96 AA; 10597 MW; AE4524418CD26F7C CRC64;

Query Match 77.3%; Score 58; DB 1; Length 96;
Best Local Similarity 75.0%; Pred. No. 0.063;
Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
| | | | |
Db 32 PKPAPWAPQKPR 43

RESULT 3
VE4 HPV31
ID VE4 HPV31 STANDARD; PRT; 102 AA.
AC P17384;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 31.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10585;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89299478; PubMed=2545036;
RA Goldsborough M.D., Disilvestre D., Temple G.F., Lorincz A.T.;
RT "Nucleotide sequence of human papillomavirus type 31: a cervical
RT neoplasia-associated virus.";
RL Virology 171:306-311(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J04353; AAA46949.1; -.
DR PIR; E32444; W4WL31.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.
SQ SEQUENCE 102 AA; 11284 MW; 04E3C9E8A5C5CA6C CRC64;

```

```

Query Match 69.3%; Score 52; DB 1; Length 102;
Best Local Similarity 80.0%; Pred. No. 0.48;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKK 10
| | | | |
Db 33 PKPAPWAPVK 42

RESULT 4
CTH1 YEAST
ID CTH1 YEAST STANDARD; PRT; 325 AA.
AC P47976;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Zinc finger protein CTH1.
GN CTH1 OR YDR151C OR YD8358.07C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Thompson M.J., Blackshear P.J.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Murphy L., Richards C., Harris D., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: Contains 2 C3H1-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L42133; AAB39897.1; -.
DR EMBL; Z50046; CAB90373.1; -.
DR PIR; S57977; S57977.
DR GerMOnline; 140642; -.
DR SGD; S0002558; CTH1.
DR InterPro; IPR000571; Znf_CCCH.
DR Pfam; PF00642; zf-CCCH; 2.
DR SMART; SM00356; Znf_C3H1; 2.
KW Nuclear protein; Repeat; Metal-binding; Zinc-finger; DNA-binding.
FT ZN FING 210 229 C3H1-TYPE 1.
FT ZN FING 248 267 C3H1-TYPE 2.
FT FT CONFLICT 142 143 RV -> EI (IN REF. 2).
SQ SEQUENCE 325 AA; 36785 MW; F1D3F46F8BF89DB CRC64;

Query Match 65.3%; Score 49; DB 1; Length 325;
Best Local Similarity 70.0%; Pred. No. 3.7;
Matches 7; Conservative 1; Mismatches 1; Indels 2; Gaps 0;

Qy 1 PKPSPWAPKK 10
| | | | |
Db 79 PRPSFWLPFK 88

RESULT 5
DCHS ENTAE
ID DCHS ENTAE STANDARD; PRT; 377 AA.
AC P28577;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)

```


10-OCT-2003 (Rel. 42, Last annotation update)
HDC. Histidine decarboxylase (EC 4.1.1.22) (HDC).
OS Enterobacter aerogenes (Aerobacter aerogenes).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Enterobacter.
OX NCBI_TaxID=548;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91236707; PubMed=2033044;
RA Kanath A.V., Vaaler G.L., Snell E.E.;
RT "Pyridoxal phosphate-dependent histidine decarboxylases. Cloning,
RT sequencing, and expression of genes from Klebsiella planticola and
RT Enterobacter aerogenes and properties of the overexpressed enzymes.";
RL J. Biol. Chem. 266:9432-9437(1991).
CC -!- CATALYTIC ACTIVITY: L-histidine = histamine + CO(2).
CC -!- COFACTOR: Pyridoxal phosphate.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SIMILARITY: Belongs to the group II decarboxylase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M62745; AAA24802.1; -;
DR PIR; A40004; A40004.
DR HAMAP; MF_00609; -; 1.
DR InterPro; IPR002129; Pyridoxal dec.
DR Pfam; PF00282; Pyridoxal dec; 1.
DR PROSITE; PS00392; DDC_GAD_HDC_YDC; 1.
KW Lyase; Decarboxylase; Pyridoxal phosphate.
FT INIT MET 0 BY SIMILARITY.
FT BINDING 232 232 PYRIDOXAL PHOSPHATE (POTENTIAL).
SQ SEQUENCE 377 AA; 42303 MW; 4C7A3334ACAD7D6AE CRC64;

Query Match 62.7%; Score 47; DB 1; Length 377;
Best Local Similarity 72.7%; Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PKPSPWAPKXH 11
|||||
DB 328 PKPSEWVWKXH 338

RESULT 6
VE4_HPV51
ID VE4_HPV51 STANDARD; PRT; 87 AA.
AC P26548;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 51.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10595;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91303675; PubMed=1649326;
RA Lungu O., Crum C.P., Silverstein S.J.;
RT "Biologic properties and nucleotide sequence analysis of human
RT papillomavirus type 51.";
RL J. Virol. 65:4216-4225(1991).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; M62877; -; NOT_ANNOTATED_CDS.
DR PIR; C40415; W4WL51.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
KW Early protein.

SQ SEQUENCE 87 AA; 9941 MW; 5F3DC38F86BF3990 CRC64;
Query Match 60.0%; Score 45; DB 1; Length 87;
Best Local Similarity 66.7%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PKPSPWAPKXH 12
|||||
DB 27 PLPPAWAPKXP 38

RESULT 7
Y514_HUMAN
ID Y514_HUMAN STANDARD; PRT; 461 AA.
AC O60269;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein KIAA0514.
GN KIAA0514.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Borak S.A., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallarón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB011086; BAA25440.1; -.
CC EMBL; BC011672; AAL11672.1; -.
CC Hypothetical protein.
CC SQ SEQUENCE 461 AA; 47657 MW; C80083C4F56CEBD CRC64;
CC -----
CC Query Match 60.0%; Score 45; DB 1; Length 461;
CC Best Local Similarity 58.3%; Pred No. 19;
CC Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
CC -----
Cc 1 PKPSFWAPKKH 12
Cc 6 PEPGFWAPLSR 17
Cc -----
Cc RESULT 8
Cc RIM1 RAT
Cc ID RIM1 RAT STANDARD; PRT; 1615 AA.
Cc AC Q9JIR4; O35168;
Cc DT 10-OCT-2003 (Rel. 42, Created)
Cc DT 10-OCT-2003 (Rel. 42, Last sequence update)
Cc DT 10-OCT-2003 (Rel. 42, Last annotation update)
Cc DE Regulating synaptic membrane exocytosis protein 1 (Rab3-interacting
Cc DE molecule 1) (RIM 1).
Cc GN RIMS1 OR RIM1.
Cc OS Rattus norvegicus (Rat).
Cc OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Cc OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Rattus.
Cc OX NCBI_TaxID=10116;
Cc RN [1]
Cc RP SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RAB3A AND RAB3C.
Cc RC TISSUE=Brain;
Cc RX MEDLINE=97394473; PubMed=9252191;
Cc RA Wang Y., Sugita S., Suedhof T.C.;
Cc RT "The RIM/NIM family of neuronal C2 domain proteins. Interactions with
Cc RT Rab3 and a new class of Src homology 3 domain proteins.";
Cc RL J. Biol. Chem. 275:20033-20044 (2000).
Cc CC -!- FUNCTION: Rab effector involved in exocytosis. May act as scaffold
Cc CC protein.
Cc CC -!- SUBUNIT: Binds SNAP25, SYT1 and CACNA1B. Interaction with SYT1 is
Cc CC enhanced by calcium ions. Interaction with SNAP25 is weaker in the
Cc CC presence of calcium ions. Binds RAB3A, RAB3B and RAB3D that have
Cc CC been activated by GTP-binding. Binds UNC13 (By similarity).
Cc CC -!- SUBCELLULAR LOCATION: Peripheral membrane protein associated with
Cc CC plasma membranes from synaptic junctions. Not detected in synaptic
Cc CC vesicles. Detected in presynaptic nerve terminals close to the
Cc CC active zone. Detected in synaptic ribbons of ribbon synapses of
Cc CC retinal photoreceptor cells.
Cc CC -!- ALTERNATIVE PRODUCTS:
Cc CC Event-Alternative splicing; Named isoforms=2;
Cc CC Name=1; Synonyms=Rim1B;
Cc CC IsoId=Q9JIR4-1; Sequences=Displayed;
Cc CC Name=2;
Cc CC IsoId=Q9JIR4-2; Sequence=VSP_008172;
Cc CC -!- TISSUE SPECIFICITY: Highly expressed in hippocampus, brain cortex,
Cc CC cerebellum and olfactory bulb. Detected at lower levels in
Cc CC midbrain, hindbrain and spinal cord. Detected retina and in spinal
Cc CC cord motor neurons.
Cc CC -!- SIMILARITY: Contains 2 C2 domains.
Cc CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.
Cc CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
Cc CC -!- SIMILARITY: Contains 1 Rab-binding (RABBD) domain.
Cc CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
Cc EMBL; AF007836; AAB66703.1; -.
Cc EMBL; AF199333; AAF81655.1; -.
Cc PIR; T03301; T03301.
Cc HSSP; P21707; IRSY.
Cc InterPro; IPR000008; C2.
Cc InterPro; IPR008973; C2 CalB.
Cc InterPro; IPR001478; PDZ.
Cc InterPro; IPR003315; RPH3A effector.
Cc InterPro; IPR000306; Znf_FYVE.
Cc Pfam; PF00168; C2; 2.
Cc Pfam; PF00595; PDZ; 1.
Cc Pfam; PF02318; RPH3A effector; 1.
Cc SMART; SM00239; C2; 2.
Cc SMART; SM00228; PDZ; 1.
Cc PROSITE; PS50004; C2 DOMAIN_2; 2.
Cc PROSITE; PS50106; PDZ; 1.
Cc PROSITE; PS50916; RABBD; 1.
Cc PROSITE; PS50178; ZF FYVE; 1.
Cc Metal-binding; Zinc; Repeat; Zinc-finger; Alternative splicing.
Cc FT DOMAIN 22 205 RAB-BINDING.
Cc FT ZN_FING 133 193 FYVE-TYPE.
Cc FT DOMAIN 422 455 ALA-RICH.
Cc FT DOMAIN 444 483 PRO-RICH.
Cc FT DOMAIN 619 705 PDZ.
Cc FT DOMAIN 758 864 C2 DOMAIN 1.
Cc FT DOMAIN 1259 1301 SER-RICH.
Cc FT DOMAIN 1461 1563 C2 DOMAIN 2.
Cc FT VARSPLIC 1107 1168 Missing (in isoform 2).
Cc FT /FTId=VSP_008172.
Cc SQ SEQUENCE 1615 AA; 179654 MW; 80E76F74BF35FB7E CRC64;
Cc -----
Cc Query Match 60.0%; Score 45; DB 1; Length 1615;
Cc Best Local Similarity 54.5%; Pred No. 60;
Cc Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Cc -----
Cc QY 1 PKPSFWAPKKH 11
Cc DB 892 PQSPFMPRRH 902
Cc -----
Cc RESULT 9
Cc RIM1 HUMAN
Cc ID RIM1 HUMAN STANDARD; PRT; 1692 AA.
Cc AC Q86UR5; O15048; Q8TDY9; Q8TDS5; Q9HBA1; Q9HBA2; Q9HBA3; Q9HBA4;
Cc AC Q9HBA5; Q9HBA6;
Cc DT 10-OCT-2003 (Rel. 42, Created)
Cc DT 10-OCT-2003 (Rel. 42, Last sequence update)
Cc DT 15-MAR-2004 (Rel. 43, Last annotation update)
Cc DE Regulating synaptic membrane exocytosis protein 1 (Rab3-interacting
Cc DE molecule 1) (RIM 1).
Cc GN RIMS1 OR RIM1 OR KIAA0340.
Cc OS Homo sapiens (Human).
Cc OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Cc OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Cc OX NCBI_TaxID=9606;
Cc RN [1]
Cc RP SEQUENCE FROM N.A. (ISOFORM 1), ALTERNATIVE SPLICING, AND VARIANT
Cc RP COR07 H1S-820.
Cc RX MEDLINE=2254706; PubMed=12659814;
Cc RA Johnson S., Halford S., Morris A.G., Patel R.J., Wilkie S.E.,
Cc RA Hardcastle A.J., Moore A.T., Zhang K., Hunt D.M.;
Cc RT "Genomic organisation and alternative splicing of human RIM1, a gene
Cc RT implicated in autosomal dominant cone-rod dystrophy (CORD7).";
Cc RL Genomics 81:304-314 (2003).
Cc RN [2]

```

SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RP TISSUE=Neuroblastoma;
 RA Aoyama M., Asai K., Shishikura T., Ohira M., Inuzuka H., Morohashi A.,
 RA Kato T., Nakagawara A.;
 RT "Identification of the alternative form of human RIM.";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani N., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 RN [4]
 RP SEQUENCE OF 776-1692 FROM N.A. (ISOFORMS 3; 4; 5; 6; 7 AND 8),
 RP MUTAGENESIS OF 796-LYS-LYS-797 AND 1591-ARG-ARG-1592, AND INTERACTION
 RP WITH SNAP25; SYT1 AND CACNA1B.
 RC TISSUE=Brain;
 RX MEDLINE=21413925; PubMed=11438518;
 RA Coppola T., Magnin-Luethi S., Perret-Menoud V., Gattesco S.,
 RA Schiavo G., Regazzi R.;
 RT "Direct interaction of the Rab3 effector RIM with Ca2+ channels, SNAP-
 RT 25, and synaptotagmin.";
 RL J. Biol. Chem. 276:32756-32762(2001).
 RN [5]
 RP FUNCTION: Rab effector involved in exocytosis. May act as scaffold
 CC protein.
 CC -! SUBUNIT: Binds RAB3A, RAB3B and RAB3D that have been activated by
 CC GTP-binding. Binds UNC13 (By similarity). Binds SNAP25, SYT1 and
 CC CACNA1B. Interaction with SYT1 is enhanced by calcium ions.
 CC Interaction with SNAP25 is weaker in the presence of calcium ions.
 CC -! ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=8;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;
 CC IsoId=Q86UR5-1; Sequence=Displayed;
 CC Name=2; Synonyms=RIM short form;
 CC IsoId=Q86UR5-2; Sequence=VSP_008160, VSP_008165;
 CC Name=3; Synonyms=RIM long form, Rab3 interacting protein variant
 CC 2;
 CC IsoId=Q86UR5-3; Sequence=VSP_008160, VSP_008164, VSP_008167,
 CC VSP_008171;
 CC Name=4; Synonyms=Rab3 interacting protein variant 1;
 CC IsoId=Q86UR5-4; Sequence=VSP_008166, VSP_008167;
 CC Name=5; Synonyms=Rab3 interacting protein variant 3;
 CC IsoId=Q86UR5-5; Sequence=VSP_008161, VSP_008163, VSP_008170;
 CC Name=6; Synonyms=Rab3 interacting protein variant 4;
 CC IsoId=Q86UR5-6; Sequence=VSP_008164, VSP_008168, VSP_008169;
 CC Name=7; Synonyms=Rab3 interacting protein variant 5;
 CC IsoId=Q86UR5-7; Sequence=VSP_008161, VSP_008164, VSP_008167,
 CC VSP_008169;
 CC Name=8; Synonyms=Rab3 interacting protein variant 6;
 CC IsoId=Q86UR5-8; Sequence=VSP_008161, VSP_008162, VSP_008169;
 CC -! TISSUE SPECIFICITY: Detected in brain and retina.
 CC -! DISEASE: Defects in RIMS1 are a cause of autosomal dominant cone-
 CC rod dystrophy (CORD7) [MIM:603649]. CORD7 is characterized by
 CC early loss of visual acuity and color vision, followed by night
 CC blindness and peripheral visual field loss. The onset of reduced
 CC color vision and visual acuity varies between the ages of 20 and
 CC 40 years.
 CC -! SIMILARITY: Contains 2 C2 domains.
 CC -! SIMILARITY: Contains 1 FIVE-type zinc finger.
 CC -! SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -! SIMILARITY: Contains 1 Rab-binding (RABBD) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.
 CC
 CC EMBL; AY190519; AAC38848.1; -
 DR EMBL; AB045726; BAB87121.1; -
 DR EMBL; AB051866; BAB87242.1; -
 DR EMBL; AB002338; BAA20798.1; ALT_INIT.
 DR EMBL; AF263305; AAG23162.1; -
 DR EMBL; AF263306; AAG23163.1; -
 DR EMBL; AF263307; AAG23164.1; -
 DR EMBL; AF263308; AAG23165.1; -
 DR EMBL; AF263309; AAG23166.1; -
 DR EMBL; AF263310; AAG23167.1; -
 DR HSP; F21707; IRSY.
 DR Genew; HGNC:17282; RIMS1.
 DR MIM; 606629; -
 DR MIM; 603649; -
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR008973; C2_CalB.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR003315; RPH3A_effector.
 DR InterPro; IPR000306; Znf_FYVE.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF02318; RPH3A_effector; 1.
 DR SMART; SM00228; PDZ; 1.
 DR SMART; SM00229; C2; 2.
 DR PROSITE; PS00004; C2_DOMAIN_2; 2.
 DR PROSITE; PS0106; PDZ; 1.
 DR PROSITE; PS0916; RABBD; 1.
 DR PROSITE; PS0178; ZF_FYVE; 1.
 KW Metal-binding; Zinc; Repeat; Zinc-finger; Alternative splicing;
 KW Vision; Disease mutation.
 FT DOMAIN 22 182 RAB-BINDING.
 FT ZN_FING 110 170 FYVE-TYPE.
 FT DOMAIN 605 691 PDZ.
 FT DOMAIN 744 850 C2 DOMAIN 1.
 FT DOMAIN 1538 1640 C2 DOMAIN 2.
 FT DOMAIN 1336 1402 SER-RICH.
 FT VARSPPLIC 484 494 Missing (in isoform 2 and isoform 3).
 FT VARSPPLIC 924 924 Missing (in isoform 5, isoform 7 and
 FT isoform 8).
 FT VARSPPLIC 1018 1245 Missing (in isoform 8).
 FT VARSPPLIC 1038 1244 Missing (in isoform 5).
 FT VARSPPLIC 1039 1102 Missing (in isoform 3, isoform 6 and
 FT isoform 7).
 FT VARSPPLIC 1040 1692 Missing (in isoform 2).
 FT VARSPPLIC 1065 1102 Missing (in isoform 4).
 FT VARSPPLIC 1133 1245 Missing (in isoform 3, isoform 4 and
 FT isoform 7).
 FT VARSPPLIC 1161 1245 Missing (in isoform 6).
 FT VARSPPLIC 1284 1455 Missing (in isoform 6, isoform 7 and
 FT isoform 8).
 FT VARSPPLIC 1377 1385 Missing (in isoform 5).
 FT VARSPPLIC 1540 1573 Missing (in isoform 3).
 FT VARIANT 820 820 R -> H (in CORD7).
 FT MUTAGEN 796 797 Missing (in isoform 6).
 FT MUTAGEN 1591 1592 Missing (in isoform 6, isoform 7 and
 FT isoform 8).
 FT SEQUENCE 1692 AA; 189071 MW; 0A96642DC832C15E CRC64;
 SQ

```

Query Match          60.0%; Score 45; DB 1; Length 1692;
Best Local Similarity 54.5%; Pred. No. 62;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPSPWAPKKH 11
   :|||:|:|
Db 878 PQSPFWAPRRH 888

RESULT 10
VE4 HPV18          STANDARD; PRT; 88 AA.
AC P06791;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-JAN-1988 (Rel. 06, Last annotation update)
DE Probable E4 protein.
OS Human papillomavirus type 18.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10582;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87283882; PubMed=3039146;
RA Cole S.T., Danos O.;
RT "Nucleotide sequence and comparative analysis of the human
RT papillomavirus type 18 genome. Phylogeny of papillomaviruses and
RT repeated structure of the E6 and E7 gene products.";
RL J. Mol. Biol. 193:599-608(1987).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X05015; CAA28668.1; -.
DR F01R; E26251; W4M118.
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
DR KW Early protein.
SQ SEQUENCE 88 AA; 9857 MW; D05F6200BF367B10 CRC64;

Query Match          58.7%; Score 44; DB 1; Length 88;
Best Local Similarity 60.0%; Pred. No. 5.7;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSPWAPKK 10
   :|||:|:|
Db 29 PAPCFWAPQR 38

RESULT 11
R122 SECC
ID R122 SECC          STANDARD; PRT; 174 AA.
AC Q06036;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 50S ribosomal protein L12-2, chloroplast precursor (Cl12-2).
GN Rpl12-2.
OS Secale cereale (Rye).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Triticeae; Secale.
OX NCBI_TaxID=4550;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Halo; TISSUE=Leaf;
RX MEDLINE=93192334; PubMed=8448215;

```

```

RA Schmidt M.W., Pichl L., Lepper M., Peierabend J.;
RT "Identification of the nuclear-encoded chloroplast ribosomal protein
RT L12 of the monocotyledonous plant Secale cereale and sequencing of
RT two different cDNAs with strong codon bias.";
RL Biochim. Biophys. Acta 1172:349-352(1993).
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X68340; CAA48414.1; -.
DR HSP; P02392; ICTF.
DR InterPro; IPR008932; Ribos_L12/7 olig.
DR Pfam; PF000206; Ribosomal_L12.
DR Pfam; PF00542; Ribosomal_L12; 1.
DR ProDom; PD001326; Ribosomal_L12; 1.
DR TIGRFAMs; TIGR00855; L12; 1.
KW Ribosomal protein; Chloroplast; Transit peptide; Multigene family.
FT TRANSIT 1 45
FT CHAIN 46 174
FT DOMAIN 16 24
FT DOMAIN 82 96
FT DOMAIN ALA-RICH.
SQ SEQUENCE 174 AA; 18319 MW; A6E20457356288C0 CRC64;

Query Match          58.7%; Score 44; DB 1; Length 174;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PKPSPWAPK 9
   :|||:|:|
Db 20 PSPPFWAPR 28

RESULT 12
TRMB HELPY
ID TRMB HELPY          STANDARD; PRT; 393 AA.
AC O25443;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE tRNA (guanine-N(7))-methyltransferase (EC 2.1.1.33) (tRNA(m7G46))-
DE methyltransferase).
GN HP0747.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
CC -!- FUNCTION: Catalyzes the formation of N(7)-methylguanine at
CC position 46 (m7G46) in tRNA (by similarity).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing N(7)-methylguanine.
CC -!- SIMILARITY: Belongs to the methyltransferase superfamily. Trmb

```

CC family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AE000587; AAD07796.1; -

DR PIR; C64613; C64613.

DR TIGR; HP0747; -

DR HAMAP; MF_01057; -; 1.

DR InterPro; IPR004395; Cons.hypoth91.

DR InterPro; IPR003358; Methyltransf_4.

DR InterPro; IPR000051; SAM_bind.

DR Pfam; PF02390; Methyltransf_4; 1.

DR TIGRFAMs; TIGR00091; TIGR00091; 1.

DR Transferrase; Methyltransf; tRNA processing; Complete proteome.

DR SEQUENCE 393 AA; 45774 MW; C376EF02D33F71D1 CRC64;

CC -----

Query Match 57.3%; Score 43; DB 1; Length 393;

Best Local Similarity 70.0%; Pred. No. 31;

Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PSPWAPKKHR 12

Db 197 PVPWNEKKHR 206

|||||

RESULT 13

ID TRMB_HELPJ STANDARD; PRT; 400 AA.

AC Q9ZL96;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE tRNA (guanine-N(7)-methyltransferase (EC 2.1.1.33) (tRNA(m7G46)-

DE methyltransferase).

GN JHP0684.

OS Helicobacter pylori J99 (Campylobacter pylori J99).

OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;

OC Helicobacteraceae; Helicobacter.

OX NCBI_TaxID=85963;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99120557; PubMed=9923682;

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown B.D., Doig P.C.,

RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,

RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,

RA Trust T.J.;

RT "Genomic sequence comparison of two unrelated isolates of the human

RT gastric pathogen Helicobacter pylori.";

RL Nature 397:176-180(1999).

CC -1- FUNCTION: Catalyzes the formation of N(7)-methylguanine at

CC position 46 (m7G46) in tRNA (By similarity).

CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-

CC homocysteine + tRNA containing N(7)-methylguanine.

CC -1- SIMILARITY: Belongs to the methyltransferase superfamily. TrmB

CC family.

CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AE001500; AAD06272.1; -

DR PIR; D71900; D71900.

DR HAMAP; MF_01057; -; 1.

DR InterPro; IPR004395; Cons.hypoth91.

DR InterPro; IPR003358; Methyltransf_4.

DR InterPro; IPR000051; SAM_bind.

DR Pfam; PF02390; Methyltransf_4; 1.

DR TIGRFAMs; TIGR00091; TIGR00091; 1.

DR Transferrase; Methyltransf; tRNA processing; Complete proteome.

DR SEQUENCE 400 AA; 46402 MW; 8B92B929183D3573 CRC64;

CC -----

Query Match 57.3%; Score 43; DB 1; Length 400;

Best Local Similarity 70.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 PSPWAPKKHR 12

Db 197 PVPWNEKKHR 206

|||||

RESULT 14

ID GTR2_HUMAN STANDARD; PRT; 524 AA.

AC P11168;

DT 01-JUN-1989 (Rel. 11, Created)

DT 01-JUN-1989 (Rel. 11, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Solute carrier family 2, facilitated glucose transporter, member 2

DE (Glucose transporter type 2, liver).

GN SLC2A2 OR GLUT2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND VARIANT LEU-68.

RC TISSUE=Kidney, and Liver;

RX MEDLINE=88289735; PubMed=3399500;

RA Fukumoto H., Seino S., Imura H., Sieno Y., Eddy R.L., Fukushima Y.,

RA Byers M.G., Shows T.B., Bell G.I.;

RT "Sequence, tissue distribution, and chromosomal localization of mRNA

RT encoding a human glucose transporter-like protein.";

RL Proc. Natl. Acad. Sci. U.S.A. 85:5434-5438(1988).

CC -1- FUNCTION: Facilitative glucose transporter. This isoform likely

CC mediates the bidirectional transfer of glucose across the plasma

CC membrane of hepatocytes and is responsible for uptake of glucose

CC by the beta cells; may comprise part of the glucose-sensing

CC mechanism of the beta cell. May also participate with the

CC Na(+)/glucose cotransporter in the transcellular transport of

CC glucose in the small intestine and kidney.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- TISSUE SPECIFICITY: Liver, insulin-producing beta cell, small

CC intestine and kidney.

CC -1- SIMILARITY: Belongs to the sugar transporter family. Glucose

CC transporter subfamily.

CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; J03810; AAA59514.1; -

DR PIR; A31318; A31318.

DR Genew; HGNC:11006; SLC2A2.

DR GK; P11168; -

DR MIM; 138160; -

DR GO; GO:0005887; C:integral to plasma membrane; TAS.

DR GO; GO:0005624; C:membrane fraction; TAS.

DR GO; GO:0005355; F:glucose transporter activity; TAS.

DR GO; GO:0005975; P:carbohydrate metabolism; TAS.

DR GO; GO:0015758; P:glucose transport; TAS.

DR InterPro; IPR007114; MFS.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:12:58 ; Search time 23.0769 Seconds
(without alignments)
164.069 Million cell updates/sec

Title: US-10-008-524A-168
Perfect score: 75
Sequence: 1 PKPSPWAPKKHR 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 25:*
1: sp archaea:*
2: sp bacteria:*
3: sp fungi:*
4: sp human:*
5: sp invertebrate:*
6: sp mammal:*
7: sp mbc:*
8: sp organelle:*
9: sp phage:*
10: sp plant:*
11: sp rodent:*
12: sp virus:*
13: sp vertebrate:*
14: sp unclassified:*
15: sp rvirus:*
16: sp bacteriaph:*
17: sp archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	75	100.0	95	12 Q918T2	Q918T2 human papil
2	75	100.0	95	12 Q918U0	Q918U0 human papil
3	75	100.0	95	12 Q918T7	Q918T7 human papil
4	75	100.0	95	12 Q8B5P2	Q8B5P2 human papil
5	75	100.0	95	12 Q8B5N7	Q8B5N7 human papil
6	75	100.0	95	12 Q80MM0	Q80MM0 human papil
7	52	69.3	88	12 Q993Z6	Q993Z6 human papil
8	52	69.3	467	10 Q9AY37	Q9AY37 oryza sativ
9	52	69.3	467	10 Q7XC25	Q7XC25 oryza sativ
10	51	68.0	464	12 Q91TM2	Q91TM2 tupaiia herp
11	51	68.0	1060	16 Q8ZFP0	Q8ZFP0 yersinia pe
12	49	65.3	2527	5 Q95W83	Q95W83 plasmodium
13	48	64.0	992	3 Q36025	Q36025 schizosacch
14	48	64.0	1242	12 Q9PZW8	Q9PZW8 eastern equ
15	48	64.0	1242	12 Q9PZW7	Q9PZW7 eastern equ
16	48	64.0	1242	12 Q9PZW6	Q9PZW6 eastern equ

17	48	64.0	1281	13 Q7ZTN2	Q7ZTN2 xenopus lae
18	47	62.7	115	5 Q9GSQ5	Q9GSQ5 ancylostoma
19	47	62.7	433	10 Q942I8	Q942I8 oryza sativ
20	46	61.3	79	4 Q8NEF7	Q8NEF7 homo sapien
21	46	61.3	230	16 Q49630	Q49630 mycobacteri
22	46	61.3	296	10 Q851Q3	Q851Q3 oryza sativ
23	46	61.3	504	5 Q95S42	Q95S42 drosophila
24	46	61.3	1052	5 Q8IR78	Q8IR78 drosophila
25	46	61.3	1235	5 Q9VYL6	Q9VYL6 drosophila
26	46	61.3	1235	5 Q9GV19	Q9GV19 drosophila
27	46	61.3	1257	5 Q9GV18	Q9GV18 drosophila
28	46	61.3	1257	5 Q8IR79	Q8IR79 drosophila
29	45	60.0	96	10 Q94HN5	Q94HN5 oryza sativ
30	45	60.0	96	10 Q7XGJ9	Q7XGJ9 oryza sativ
31	45	60.0	364	10 Q8LQLO	Q8LQLO oryza sativ
32	45	60.0	1095	10 Q9SM80	Q9SM80 oryza sativ
33	45	60.0	2480	5 Q86MB6	Q86MB6 plasmodium
34	44.5	59.3	1312	4 Q9NR59	Q9NR59 homo sapien
35	44.5	59.3	1312	4 Q7Z5V7	Q7Z5V7 homo sapien
36	44.5	59.3	1343	4 Q9H7N4	Q9H7N4 homo sapien
37	44	58.7	186	10 Q9FRD0	Q9FRD0 oryza sativ
38	44	58.7	888	10 Q9LST1	Q9LST1 arabidopsis
39	43	57.3	72	4 Q9UCW9	Q9UCW9 homo sapien
40	43	57.3	123	4 Q8WYV3	Q8WYV3 homo sapien
41	43	57.3	255	10 Q8S6Y1	Q8S6Y1 oryza sativ
42	43	57.3	255	10 Q7XG79	Q7XG79 oryza sativ
43	43	57.3	307	17 Q8ZT34	Q8ZT34 pyrobaculum
44	43	57.3	356	10 Q9LIP1	Q9LIP1 arabidopsis
45	43	57.3	435	11 Q91VI8	Q91VI8 mus musculu

ALIGNMENTS

RESULT 1
Q918T2 ID Q918T2 PRELIMINARY; PRT; 95 AA.
AC Q918T2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE E4 protein (Fragment).
GN E4.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
CX NCBI_TaxID=10581;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HPV16E2CC7;
RA Watts K.J., Thompson C.H., Cossart Y.E., Rose B.R.;
RT "Sequence variation and physical state of human Papillomavirus type 16
RT cervical cancer isolates from Australia and New Caledonia.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF407220; AAL01405.1; -
DR InterPro; IPR003861; Papilloma_E4.
DR Pfam; PF02711; Pap_E4; 1.
FT NON TER 1
SQ SEQUENCE 95 AA; 10542 MW; AED17903867307CE CRC64;

Query Match 100.0%; Score 75; DB 12; Length 95;
Best Local Similarity 100.0%; Pred. No. 0.00021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
Db 32 PKPSPWAPKKHR 43

RESULT 2
Q918U0 ID Q918U0 PRELIMINARY; PRT; 95 AA.
AC Q918U0;

DT	01-DEC-2001	(TrEMBLrel. 19, Created)
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)
DT	01-MAR-2002	(TrEMBLrel. 20, Last annotation update)
DE	E4	protein (Fragment).
OS	Human papillomavirus type 16.	
OC	Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;	
OC	Papillomavirus.	
NCBI_TaxID=10581;		
GN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=HPV16E2CC4;	
RA	Watts K.J., Thompson C.H., Cossart Y.E., Rose B.R.;	
RT	"Sequence variation and physical state of human Papillomavirus type 16	
RL	serical cancer isolates from Australia and New Caledonia.";	
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF407217; AAL01396.1; -.	
DR	InterPro; IPR003861; Papilloma_E4.	
DR	Pfam; PF02711; Pap_E4; 1.	
FT	NON_TER	1
SQ	SEQUENCE	95 AA; 10608 MW; AED4269D05E307CE CRC64;
Query Match		
Best Local Similarity		
Matches		
12; Conservative		
0; Mismatches		
0; Indels		
0; Gaps		
0;		
Qy	1	PKPSPWAPKKHR 12
Db	32	PKPSPWAPKKHR 43
RESULT 3		
Q918T7		
ID	Q918T7	PRELIMINARY; PRT; 95 AA.
AC	Q918T7;	
DT	01-DEC-2001	(TrEMBLrel. 19, Created)
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)
DT	01-MAR-2002	(TrEMBLrel. 20, Last annotation update)
DE	E4	protein (Fragment).
GN	E4.	
OS	Human papillomavirus type 16.	
OC	Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;	
OC	Papillomavirus.	
NCBI_TaxID=10581;		
GN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=HPV16E2CC5;	
RA	Watts K.J., Thompson C.H., Cossart Y.E., Rose B.R.;	
RT	"Sequence variation and physical state of human Papillomavirus type 16	
RL	serical cancer isolates from Australia and New Caledonia.";	
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF407218; AAL01399.1; -.	
DR	InterPro; IPR003861; Papilloma_E4.	
DR	Pfam; PF02711; Pap_E4; 1.	
FT	NON_TER	1
SQ	SEQUENCE	95 AA; 10580 MW; BE0E469D177307CE CRC64;
Query Match		
Best Local Similarity		
Matches		
12; Conservative		
0; Mismatches		
0; Indels		
0; Gaps		
0;		
Qy	1	PKPSPWAPKKHR 12
Db	32	PKPSPWAPKKHR 43
RESULT 4		
Q8B5P2		
ID	Q8B5P2	PRELIMINARY; PRT; 95 AA.
AC	Q8B5P2;	
DT	01-MAR-2003	(TrEMBLrel. 23, Created)
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)

OC Papillomavirus.
 RN [1]_TaxID=10561;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Asian-American variant;
 RA Terai M., Burk R.D.;
 RT "Human papillomavirus type 16 Asian-American variant."
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF402678; AAO85412.1; -;
 DR InterPro: IPR003861; Papilloma_E4.
 DR Pfam: PF02711; Pap_E4; 1.
 FT NON TER 1
 SQ SEQUENCE 95 AA; 10542 MW; AED17903867307CE CRC64;

Query Match 100.0%; Score 75; DB 12; Length 95;
 Best Local Similarity 100.0%; Pred. No. 0.00021;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 32 PKPSPWAPKKHR 43

RESULT 7
 Q993Z6 PRELIMINARY; PRT; 88 AA.
 AC Q993Z6;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE Putative E4 protein.
 GN E4.
 OS Human papillomavirus type 82.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OC NCBI_TaxID=129724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Terai M., Burk R.D.;
 RT "Cervical HPV in Evolution; Genomic Sequence of IS39/AE2, a Subtype of Oncogenic HPV 82 (W13B).";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF293961; AAK28453.1; -;
 DR InterPro: IPR003861; Papilloma_E4.
 DR Pfam: PF02711; Pap_E4; 1.
 SQ SEQUENCE 88 AA; 10084 MW; 6752D8CF3A9475D7 CRC64;

Query Match 69.3%; Score 52; DB 12; Length 88;
 Best Local Similarity 75.0%; Pred. No. 0.75;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 27 PIPPPWAPKKPR 38

RESULT 8
 Q9AY37 PRELIMINARY; PRT; 467 AA.
 AC Q9AY37;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative phosphatidylinositol-4-phosphate 5-kinase.
 GN OSUNBA0027P10.22.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OC NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;

RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Hsiao J.,
 RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E., Craven B.,
 RA Khalak H., Feldblyum T.V., Quackenbush J., White O., Salzberg S.L.,
 RA Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBa0027P10 genomic sequence."
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC084763; AAG60194.1; -;
 DR Gramene; Q9AY37; -;
 DR GO: GO:0016301; F-kinase activity; IEA.
 DR InterPro: IPR003409; MORN.
 DR Pfam: PF02493; MORN; 7.
 DR SMART; SM00698; MORN; 7.
 KW Kinase.
 SQ SEQUENCE 467 AA; 50906 MW; 8FB85BCC9B980D73 CRC64;

Query Match 69.3%; Score 52; DB 10; Length 467;
 Best Local Similarity 66.7%; Pred. No. 4;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 75 PPSPPAPSRHR 86

RESULT 9
 Q7XC25 PRELIMINARY; PRT; 467 AA.
 AC Q7XC25;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative phosphatidylinositol-4-phosphate 5-kinase.
 GN OSJNBa0027P10.22.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OC NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=cv. Nipponbare;
 RC The Rice Chromosome 10 Sequencing Consortium;
 RT "In-depth view of structure, activity, and evolution of rice chromosome 10."
 RL Science 300:1566-1569(2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE017120; AAP5050.1; -;
 KW Kinase.
 SQ SEQUENCE 467 AA; 50906 MW; 8FB85BCC9B980D73 CRC64;

Query Match 69.3%; Score 52; DB 10; Length 467;
 Best Local Similarity 66.7%; Pred. No. 4;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 |||||
 Db 75 PPSPPAPSRHR 86

RESULT 10
 Q91TM2 PRELIMINARY; PRT; 464 AA.
 AC Q91TM2;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE T74.
 OS Tupaia herpesvirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Betaherpesvirinae.
 OX NCBI_TaxID=10397;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2;
 RX MEDLINE=21211637; PubMed=11312357;
 RA Bahr U., Darai G.;
 RT "Analysis and Characterization of the Complete Genome of Tupaia (Tree Shrew) Herpesvirus.";
 RL J. Virol. 75:4854-4870 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=2;
 RA Darai G., Bahr U.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF281817; AAK57119.1; -;
 SQ SEQUENCE 464 AA; 51193 MW; 4BB7313EA2C2BD16 CRC64;

Query Match 68.0%; Score 51; DB 12; Length 464;
 Best Local Similarity 66.7%; Pred. No. 5.6;
 Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 | | | | | | | |
 Db 402 PPPSPWRPPLHR 413

RESULT 11
 ID Q8ZFP0 PRELIMINARY; PRT; 1060 AA.
 AC Q8ZFP0;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Beta-galactosidase (EC 3.2.1.23) (Beta-D-galactosidase).
 GN LACZ OR YPO1654 OR Y1817.
 OS *Yersinia pestis*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Yersinia*.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.B., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Hollroyd S., Jagels K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of *Yersinia pestis*, the causative agent of plague.";
 RL Nature 413:523-527 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM5 / Biovar Mediaevalis;
 RX MEDLINE=22137863; PubMed=12142430;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of *Yersinia pestis* KIM.";
 RL J. Bacteriol. 184:4601-4611 (2002).
 DR EMBL; AJ414149; CAC90476.1; -;
 DR EMBL; AE013784; AAM85385.1; -;
 DR PIR; A10201; A10201.
 DR GO; GO:0009341; C:beta-galactosidase complex; IEA.
 DR GO; GO:0004565; F:beta-galactosidase activity; IEA.
 DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR008979; Gal bind like.
 DR InterPro; IPR006101; Glyco_hydro_2.

DR InterPro; IPR006102; Glyco_hydro_2Ig.
 DR InterPro; IPR006104; Glyco_hydro_2SB.
 DR InterPro; IPR006103; Glyco_hydro_2TIM.
 DR InterPro; IPR004200; Glyco_hydro_42C.
 DR InterPro; IPR004199; Glyco_hydro_42N.
 DR Pfam; PF02930; Bgal_small_C; 1.
 DR Pfam; PF02929; Bgal_small_N; 1.
 DR Pfam; PF00703; Glyco_hydro_2; 1.
 DR Pfam; PF02836; Glyco_hydro_2_C; 1.
 DR Pfam; PF02837; Glyco_hydro_2_N; 1.
 DR PRINTS; PR00132; GLHYDLASE2.
 DR PROSITE; PS00719; GLYCOSYL HYDROL_F2_1; 1.
 DR PROSITE; PS00608; GLYCOSYL HYDROL_F2_2; 1.
 KW Glycosidase; Hydrolase; Complete proteome.
 SQ SEQUENCE 1060 AA; 122625 MW; 1142D7FC68E46697 CRC64;

Query Match 68.0%; Score 51; DB 16; Length 1060;
 Best Local Similarity 58.3%; Pred. No. 13;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 | | | | | | | |
 Db 723 PKETPWSPAQHR 734

RESULT 12
 ID Q95W83 PRELIMINARY; PRT; 2527 AA.
 AC Q95W83;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Erythrocyte membrane protein 1 (fragment).
 GN VAK1.
 OS *Plasmodium falciparum*.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21442075; PubMed=11557894;
 RA Flick K., Scholander C., Chen Q., Fernandez V., Pouvelle B., Gysin J.,
 RA Wahlgren M.;
 RT "Role of nonimmune IgG bound to PfEMP1 in placental malaria.";
 RL Science 293:2098-2100 (2001).
 DR EMBL; AF365657; AAL12845.1; -;
 DR GO; GO:0005539; F:glycosaminoglycan binding; IEA.
 DR GO; GO:0009405; P:pathogenesis; IEA.
 DR InterPro; IPR004258; PFEMP.
 DR Pfam; PF03011; PFEMP; 2.
 FT NON_TER 2527 2527
 SQ SEQUENCE 2527 AA; 292137 MW; 59EE2766BF0425E4 CRC64;

Query Match 65.3%; Score 49; DB 5; Length 2527;
 Best Local Similarity 75.0%; Pred. No. 63;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PKPSPWAPKKHR 12
 | | | | | | | |
 Db 2404 PKPGKAPKKR 2415

RESULT 13
 ID Q36025 PRELIMINARY; PRT; 992 AA.
 AC Q36025;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN SPAC4F10.13C.
 OS *Schizosaccharomyces pombe* (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;

```

OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1] | | | | |
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98980; CAB11716.1; -.
DR PIR; T38817; T38817.
DR GeneDB_Spombe; SPAC4F10.13C; -.
DR InterPro; IPR003169; GYF.
DR Pfam; PF02213; GYF; 1.
DR SMART; SM00444; GYF; 1.
DR PROSITE; PS00829; GYF; 1.
DR Hypothetical protein.
KW Hypothetical protein.
SQ SEQUENCE 992 AA; 106142 MW; 4751B990B0C39345 CRC64;

Query Match 64.0%; Score 48; DB 3; Length 992;
Best Local Similarity 69.2%; Pred. No. 35;
Matches 9; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 1 PKPSPW--APKKH 11
DB 622 PKPSPWKSLLPEKH 634

RESULT 14
Q9PZW8 PRELIMINARY; PRT; 1242 AA.
ID Q9PZW8
AC Q9PZW8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Structural polyprotein.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
OS virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11021;
RN [1] | | | | |
RP SEQUENCE FROM N.A.
RC STRAIN=BR56-BeAn5122;
RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159559; AAF04801.1; -.
DR HSSP; P03315; 1VCP.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0004352; F:serine-type endopeptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR00930; Peptidase_S3.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
KW Polyprotein.
SQ SEQUENCE 1242 AA; 137436 MW; 25A0990E69D16636 CRC64;

Query Match 64.0%; Score 48; DB 12; Length 1242;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 PKPSPWAPKKHR 12
DB 73 PKPSPAPKKKR 84

RESULT 15
Q9PZW7 PRELIMINARY; PRT; 1242 AA.
ID Q9PZW7
AC Q9PZW7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Structural polyprotein.
OS Eastern equine encephalitis virus (Eastern equine encephalomyelitis
OS virus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Alphavirus.
OX NCBI_TaxID=11021;
RN [1] | | | | |
RP SEQUENCE FROM N.A.
RC STRAIN=PA86-435731;
RA Brault A.C., Powers A.M., Kang W., Tesh R.B., Shope R.E., Weaver S.C.;
RT "Genetic and Antigenic Diversity among Eastern Equine encephalitis
RT viruses from North, Central and South America.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF159560; AAF04802.1; -.
DR HSSP; P03315; 1VCP.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0004352; F:serine-type endopeptidase activity; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002548; Alpha_E1_glycop.
DR InterPro; IPR000936; Alpha_E2_glycop.
DR InterPro; IPR002533; Alpha_E3_glycop.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR00930; Peptidase_S3.
DR Pfam; PF00944; Alpha_core; 1.
DR Pfam; PF01589; Alpha_E1_glycop; 1.
DR Pfam; PF00943; Alpha_E2_glycop; 1.
DR Pfam; PF01563; Alpha_E3_glycop; 1.
DR PRINTS; PR00798; TOGAVIRIN.
KW Polyprotein.
SQ SEQUENCE 1242 AA; 137344 MW; 2759F20F690B4A13 CRC64;

Query Match 64.0%; Score 48; DB 12; Length 1242;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 PKPSPWAPKKHR 12
DB 73 PKPSPAPKKKR 84

Search completed: May 27, 2004, 16:20:57
Job time : 24.0769 secs

```

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 27, 2004, 16:03:03 ; Search time 69.2308 Seconds
(without alignments)
102.031 Million cell updates/sec

Title: US-10-008-524A-4
Perfect score: 142
Sequence: 1 RPIPKPSEWAPKRRRLSDQDSQTP 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001s:.*
5: Geneseq2002s:.*
6: Geneseq2003as:.*
7: Geneseq2003bs:.*
8: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	131.5	92.6	26	2	AAW62277 HPV16 hyd
2	131.5	92.6	26	5	AAU75260 Hydrophil
3	110.5	77.8	30	2	AAR15564 Immunopep
4	98.5	69.4	20	2	AAR15562 Immunopep
5	82	57.7	15	2	AAW62278 HPV16 hyd
6	81	57.0	14	2	AAAR14288 Seroreact
7	75	52.8	12	2	AAW62279 HPV16 hyd
8	66	46.5	11	2	AAR14287 Seroreact
9	61.5	43.3	1053	7	ADRS4638 Human Pro
10	61.5	43.3	1553	7	ADRS4636 Rat Prote
11	61	43.0	51	4	AAU51469 Propionib
12	61	43.0	51	6	ABM47988 Propionib
13	58	40.8	985	4	ABBS9214 Drosophil
14	57	40.1	230	5	ABU05501 M. tuberc
15	56	39.4	105	4	AAU03023 Human pol
16	56	39.4	394	4	ABG16932 Novel hum
17	55.5	39.1	525	5	AAB47787 Protein e
18	54	38.0	77	4	AAU55993 Propionib
19	54	38.0	77	6	ABMS2512 Propionib
20	54	38.0	121	4	ABG19332 Novel hum
21	54	38.0	177	7	ABM74128 DNA clone
22	54	38.0	682	7	ADC37582 Human nuc
23	54	38.0	683	5	ABBS97217 Novel hum
24	54	38.0	2451	4	ABB71574 Drosophil
25	53.5	37.7	598	2	AAW99065 Human U62

26	53	37.3	288	4	ABG09463	Abg09463 Novel hum
27	53	37.3	356	4	AAm38780	AAm38780 Human pol
28	53	37.3	356	7	ADD18690	ADD18690 Human dis
29	53	37.3	382	3	AAAG50815	AAAG50815 Arabidops
30	53	37.3	382	3	AAAG17584	AAAG17584 Arabidops
31	53	37.3	387	3	AAAG17583	AAAG17583 Arabidops
32	53	37.3	387	3	AAAG50814	AAAG50814 Arabidops
33	53	37.3	387	3	AAAG50814	AAAG50814 Arabidops
34	53	37.3	495	4	AAm40566	AAm40566 Human pol
35	52.5	37.0	394	5	AAO17206	AAO17206 Novel hum
36	52.5	37.0	394	5	ABG64749	ABG64749 Human sec
37	52.5	37.0	586	3	AAAG30917	AAAG30917 Arabidops
38	52.5	37.0	615	3	AAAG30916	AAAG30916 Arabidops
39	52.5	37.0	684	3	AAAG30915	AAAG30915 Arabidops
40	52.5	37.0	1590	4	AAAB73488	AAAB73488 Mouse Rim
41	52	36.6	112	3	AAAG02743	AAAG02743 Human sec
42	52	36.6	184	3	AAAB37424	AAAB37424 Human sec
43	52	36.6	184	3	AAAB36111	AAAB36111 Human MTC
44	52	36.6	184	4	AAAB47140	AAAB47140 CDIFF-21,
45	52	36.6	184	5	AAO15355	AAO15355 Human sig

ALIGNMENTS

RESULT 1
AAW62277
ID AAW62277 standard; peptide; 26 AA.
XX
AC AAW62277;
XX
DT 24-SEP-1998 (first entry)
XX
DE HPV16 hydrophilic region peptide #1.
XX
KW HPV16; human papilloma virus; epithelial tumour; cervical cancer;
KW precancerous cervical lesion; screening; detection; infection; cervix;
KW HPV E4.
XX
OS Human papillomavirus.
XX
PN WO9825145-A1.
XX
PD 11-JUN-1998.
XX
PF 03-DEC-1997; 97WO-GB003321.
XX
PR 03-DEC-1996; 96GB-00025142.
PR 05-SEP-1997; 97GB-00018745.
(MEDI-) MEDICAL RES COUNCIL.
Doorbar J;
WPI; 1998-333497/29.
Detecting papilloma virus infection using molecule binding to E4 protein
- useful, e.g. in screening for pre-cancerous cervical lesions and to
determine type(s) of human papilloma virus infecting human patients.
Claim 9; Page 37; 52pp; English.

A new method has been developed for detecting a papilloma virus infection in an organism. The method comprises: (i) obtaining a sample of cells from the potential infection site; (ii) contacting the cells with a molecule binding specifically to papilloma virus E4 protein, and (iii) monitoring the binding. The method is useful to detect papilloma virus infections in organisms (especially mammals) and especially HPV infections (e.g. with HPV16, 18, 33, 35, 45, 51, 56, 58 or 61) in humans. Papilloma viruses cause epithelial tumours in humans varying in severity depending on the infection site and HPV type involved. The method is particularly useful to determine papilloma infection in the mammalian cervix and especially to screen for pre-cancerous cervical lesions in

CC humans, since over 90% of cervical carcinoma patients show cervical HPV
 CC infection. It is also useful to determine the type(s) of HPV infection in
 CC a patients, by using a molecule binding specifically to a subset of HPV E4
 CC proteins. This is important, since progression to malignant disease (and
 CC hence clinical prognosis) is dependent on HPV type. Molecules capable of
 CC binding E4 are also useful to target anticancer/antiviral agents capable
 CC of destroying papilloma viruses and/or papilloma virus-infected cells.
 CC The present sequence represents a specifically claimed HPV16 peptide
 CC found in the hydrophilic region
 XX
 XX Sequence 26 AA;
 SQ
 Query Match 92.6%; Score 131.5; DB 2; Length 26;
 Best Local Similarity 96.2%; Pred. No. 6e-11;
 Matches 25; Conservative 0; Mismatches 1; Gaps 1;
 QY 1 RPIPKSPWAPKKHRL-SDQDSQTP 25
 Db 1 RPIPKSPWAPKKHRLSSDQDSQTP 26
 RESULT 2
 AAU75260
 ID AAU75260 standard; peptide; 26 AA.
 AC
 AC AAU75260;
 XX
 XX 21-MAY-2002 (first entry)
 DT
 DE Hydrophilic region of HPV16 E4 protein.
 XX
 XX Papilloma virus associated antigen; cell proliferation marker;
 KW cervical malignancy; human papillomavirus infection; HPV; lesion;
 KW cellular abnormality; cellular proliferation; cellular growth; dysplasia;
 KW neoplasia; cancer; papilloma smear test; E4 protein.
 XX
 OS Human papillomavirus type 16.
 XX
 XX WO200208764-A1.
 FN
 XX 31-JAN-2002.
 PD
 XX 16-MAR-2001; 2001WO-GB001176.
 PF
 XX 24-JUL-2000; 2000GB-00018140.
 PR
 XX (MEDI-) MEDICAL RES COUNCIL.
 PA
 PI Doorbar J;
 PI
 XX WPI; 2002-188648/24.
 DR
 XX
 XX Detecting abnormalities e.g. abnormal cellular proliferation, in a sample
 PT from a patient comprises contacting cells with a molecule which can bind
 PT a papilloma virus associated antigen, or a cell proliferation or viral
 PT activity marker.
 XX
 PS Disclosure; Page 23; 90pp; English.
 XX
 XX The present invention relates to a method for detecting abnormalities in
 CC a sample from a patient. The method comprises contacting a sample of the
 CC patient's cells with two or more molecules, where at least one molecule
 CC is capable of binding a papilloma virus associated antigen, and at least
 CC one molecule is capable of binding a cell proliferation marker. The
 CC method is useful for simultaneously screening for abnormalities which
 CC indicate or can lead to cervical malignancy, for human papillomavirus
 CC (HPV) infections, and precursor lesions or other conditions which occur
 CC with cervical malignancy. The method is also useful for assessing the
 CC risk associated with cellular abnormality in a patient sample, and for
 CC determining, assessing or diagnosing the presence or absence of abnormal
 CC cellular proliferation, cellular growth abnormality, dysplasia,
 CC neoplasia, or a pre-cancerous or cancerous state in a tissue. The new
 CC method is much simpler, and yields more information more quickly than

CC conventional papilloma smear testing programmes. Compared with previous
 CC methods of screening, the new method has reduced chances of false
 CC negatives occurring, requires fewer samples to gain the same amount of
 CC information, and alleviates the need for repeated or further testing. The
 CC present sequence representing the hydrophilic region of the HPV16 E4
 CC protein binds antibody molecules
 XX
 XX Sequence 26 AA;
 SQ
 Query Match 92.6%; Score 131.5; DB 5; Length 26;
 Best Local Similarity 96.2%; Pred. No. 6e-11;
 Matches 25; Conservative 0; Mismatches 1; Gaps 1;
 QY 1 RPIPKSPWAPKKHRL-SDQDSQTP 25
 Db 1 RPIPKSPWAPKKHRLSSDQDSQTP 26
 RESULT 3
 AAR15564
 ID AAR15564 standard; protein; 30 AA.
 XX
 AC AAR15564;
 AC
 XX 02-MAR-1992 (first entry)
 DT
 XX
 DE Immunopeptide #4 derived from HPV16 E4 peptide.
 XX
 XX cervical cancer; cervical intraepithelial neoplasia; CIN; wart;
 KW squamous cell carcinoma; ELISA; HPV 16.
 KW
 OS Synthetic.
 OS
 XX WO9118294-A.
 PN
 XX 28-NOV-1991.
 PD
 XX 11-MAY-1990; 90SE-00001705.
 PF
 XX 11-MAY-1990; 90SE-00001705.
 PR
 XX (MEDS-) MEDSCAND AB.
 PA
 PI Dillner J, Dillner L, Cheng HM;
 PI
 XX WPI; 1991-369390/50.
 DR
 XX Diagnosis of human papilloma virus infection and pv-carrying tumours -
 PT using synthetic peptide(s) to detect virus specific antigen-antibody
 PT complexes by immunoassay.
 PT
 XX Disclosure; Page 38; 72pp; English.
 PS
 XX This is one of a large number of peptides which have been synthesised on
 CC the basis of the amino acid sequences for the E2, E4, E7, L1 or L2
 CC proteins of HPV 1, 5, 6, 8, 11, 16, 18, 31 and 33. The selection of
 CC peptide sequences was based on the assumption that an immunoreactive
 CC region might be situated in the same relative region of a protein from
 CC different HPV types. The peptides were used in diagnostic immunoassays to
 CC detect HPV-infection. See AAR15523-R15601
 CC
 XX
 SQ Sequence 30 AA;
 Query Match 77.8%; Score 110.5; DB 2; Length 30;
 Best Local Similarity 95.5%; Pred. No. 5.2e-08;
 Matches 21; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 RPIPKSPWAPKKHRL-SDQD 21
 Db 7 RPIPKSPWAPKKHRLSSDQD 28
 RESULT 4.

```

AAR15562
ID AAR15562 standard; protein; 20 AA.
XX
AC AAR15562;
XX
DT 02-MAR-1992 (first entry)
XX
DE Immunopeptide #2 derived from HPV16 E4 peptide.
XX
KW cervical cancer; cervical intraepithelial neoplasia; CIN; wart;
KW squamous cell carcinoma; ELISA; HPV 16.
XX
OS Synthetic.
XX
PN W09118294-A.
XX
PD 28-NOV-1991.
XX
PF 11-MAY-1990; 90SE-00001705.
XX
PR 11-MAY-1990; 90SE-00001705.
XX
PA (MEDS-) MEDSCAND AB.
XX
PI Dillner J, Dillner L, Cheng HM;
XX
WPI; 1991-369390/50.
XX
DR
XX
PT Diagnosis of human papilloma virus infection and PV-carrying tumours -
PT using synthetic peptide(s) to detect virus specific antigen-antibody
PT complexes by immunoassay.
XX
PS Disclosure; Page 38; 72pp; English.
XX
CC This is one of a large number of peptides which have been synthesised on
CC the basis of the amino acid sequences for the E2, E4, E7, I1 or L2
CC proteins of HPV 1, 5, 6, 8, 11, 16, 18, 31 and 33. The selection of
CC peptide sequences was based on the assumption that an immunoreactive
CC region might be situated in the same relative region of a protein from
CC different HPV types. The peptides were used in diagnostic immunoassays to
CC detect HPV-infection. See AAR15523-R15601
XX
SQ Sequence 20 AA;
Query Match 69.4%; Score 98.5; DB 2; Length 20;
Best Local Similarity 95.0%; Pred. No. 1.5e-06;
Matches 19; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 3 IPKPSWAPKKHRL-SDQD 21
Db 1 IPKPSWAPKKHRLSDQD 20
RESULT 5
AAR62278
ID AAR62278 standard; peptide; 15 AA.
XX
AC AAR62278;
XX
DT 24-SEP-1998 (first entry)
XX
DE HPV16 hydrophilic region peptide #2.
XX
KW HPV16; human papilloma virus; epithelial tumour; cervical cancer;
KW precancerous cervical lesion; screening; detection; infection; cervix;
KW HPV E4.
XX
OS Human papillomavirus.
XX
PN W09825145-A1.
XX
PD 11-JUN-1998.
XX

```

```

PF 03-DEC-1997; 97WO-GB003321.
XX
PR 03-DEC-1996; 96GB-00025142.
PR 05-SEP-1997; 97GB-00018745.
XX
PA (MEDI-) MEDICAL RES COUNCIL.
XX
PI Doorbar J;
XX
WPI; 1998-333497/29.
XX
DR
XX
PT Detecting papilloma virus infection using molecule binding to E4 protein
PT - useful, e.g. in screening for pre-cancerous cervical lesions and to
PT determine type(s) of human papilloma virus infecting human patients.
XX
PS Claim 10; Page 37; 52pp; English.
XX
CC A new method has been developed for detecting a papilloma virus infection
CC in an organism. The method comprises: (i) obtaining a sample of cells
CC from the potential infection site; (ii) contacting the cells with a
CC molecule binding specifically to papilloma virus E4 protein, and (iii)
CC monitoring the binding. The method is useful to detect papilloma virus
CC infections in organisms (especially mammals) and especially HPV
CC infections (e.g. with HPV16, 18, 33, 35, 45, 51, 56, 58 or 61) in humans.
CC Papilloma viruses cause epithelial tumours in humans varying in severity
CC depending on the infection site and HPV type involved. The method is
CC particularly useful to determine papilloma infection in the mammalian
CC cervix and especially to screen for pre-cancerous cervical lesions in
CC humans, since over 90% of cervical carcinoma patients show cervical HPV
CC infection. It is also useful to determine the type(s) of HPV infection in
CC a patient, by using a molecule binding specifically to a subset of HPV E4
CC proteins. This is important, since progression to malignant disease (and
CC hence clinical prognosis) is dependent on HPV type. Molecules capable of
CC binding E4 are also useful to target anticancer/antiviral agents capable
CC of destroying papilloma viruses and/or papilloma virus-infected cells.
CC The present sequence represents a specifically claimed HPV16 peptide
CC found in the hydrophilic region
XX
SQ Sequence 15 AA;
Query Match 57.7%; Score 82; DB 2; Length 15;
Best Local Similarity 93.3%; Pred. No. 0.0002;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 RPIPKPSWAPKKHR 15
Db 1 RRIKPSWAPKKHR 15
RESULT 6
AAR14288
ID AAR14288 standard; protein; 14 AA.
XX
AC AAR14288;
XX
DT 02-JAN-1992 (first entry)
XX
DE Seroreactive epitope #2 of HPV16 protein E4.
XX
KW HPV16-dependent human disease; E6; E7; L1.
XX
OS Synthetic.
XX
PN EP451550-A.
XX
PD 16-OCT-1991.
XX
PF 19-MAR-1991; 91EP-00104197.
XX
PR 20-MAR-1990; 90EP-00105222.
XX
PA (BEHW ) BEHRINGWERKE AG.
XX

```

PI Muller M, Gissmann L;
 XX WPI; 1991-304643/42.
 DR Sero-active epitope(s) of human papilloma-virus 16 proteins - for use
 XX vaccines and diagnosis.
 PT Claim 1; Page 11; 15pp; English.
 XX
 CC This is one of two seroreactive epitopes identified from HPV16 protein
 CC E4. Peptides containing this epitope are also claimed and can be used to
 CC generate antibodies to HPV. See AAR14287-R14302 and AAQ14168-Q14171
 XX
 SQ Sequence 14 AA;
 Query Match 57.0%; Score 81; DB 2; Length 14;
 Best Local Similarity 100.0%; Pred. No. 0.00026;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 KPSPWAPKKHRLS 18
 Db 1 KPSPWAPKKHRLS 14
 RESULT 7
 AAW62279
 ID AAW62279 standard; peptide; 12 AA.
 XX
 AC AAW62279;
 XX
 DT 24-SEP-1998 (first entry)
 XX
 DE HPV16 hydrophilic region peptide #3.
 XX
 KW HPV16; human papilloma virus; epithelial tumour; cervical cancer;
 KW precancerous cervical lesion; screening; detection; infection; cervix;
 KW HPV E4.
 XX
 OS Human papillomavirus.
 XX
 FN WO9825145-A1.
 XX
 PD 11-JUN-1998.
 XX
 PF 03-DEC-1997; 97WO-GB003321.
 XX
 PR 03-DEC-1996; 96GB-00025142.
 PR 05-SEP-1997; 97GB-00018745.
 XX
 PA (MEDI-) MEDICAL RES COUNCIL.
 XX
 PI Doorbar J;
 XX
 DR WPI; 1998-333497/29.
 XX
 PT Detecting papilloma virus infection using molecule binding to E4 protein
 PT - useful, e.g. in screening for pre-cancerous cervical lesions and to
 PT determine type(s) of human papilloma virus infecting human patients.
 XX
 PS Claim 11; Page 37; 52pp; English.
 XX
 CC A new method has been developed for detecting a papilloma virus infection
 CC in an organism. The method comprises: (i) obtaining a sample of cells
 CC from the potential infection site; (ii) contacting the cells with a
 CC molecule binding specifically to papilloma virus E4 protein; and (iii)
 CC monitoring the binding. The method is useful to detect papilloma virus
 CC infections in organisms (especially mammals) and especially HPV
 CC infections (e.g. with HPV16, 18, 33, 35, 45, 51, 56, 58 or 61) in humans.
 CC Papilloma viruses cause epithelial tumours in humans varying in severity
 CC depending on the infection site and HPV type involved. The method is
 CC particularly useful to determine papilloma infection in the mammalian
 CC cervix and especially to screen for pre-cancerous cervical lesions in
 CC humans, since over 90% of cervical carcinoma patients show cervical HPV

CC infection. It is also useful to determine the type(s) of HPV infection in
 CC a patient, by using a molecule binding specifically to a subset of HPV E4
 CC proteins. This is important, since progression to malignant disease (and
 CC hence clinical prognosis) is dependent on HPV type. Molecules capable of
 CC binding E4 are also useful to target anticancer/antiviral agents capable
 CC of destroying papilloma viruses and/or papilloma virus-infected cells.
 CC The present sequence represents a specifically claimed HPV16 peptide
 XX found in the hydrophilic region
 SQ Sequence 12 AA;
 Query Match 52.8%; Score 75; DB 2; Length 12;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 KPSPWAPKKHR 15
 Db 1 KPSPWAPKKHR 12
 RESULT 8
 AAR14287
 ID AAR14287 standard; peptide; 11 AA.
 XX
 AC AAR14287;
 XX
 DT 02-JAN-1992 (first entry)
 XX
 DE Seroreactive epitope #1 of HPV16 protein E4.
 XX
 KW HPV16-dependent human disease; E6; E7; L1.
 XX
 OS Synthetic.
 XX
 FN EP451550-A.
 PD 16-OCT-1991.
 XX
 PF 19-MAR-1991; 91EP-00104197.
 XX
 PR 20-MAR-1990; 90EP-00105222.
 XX
 PA (BEHW) BEHRINGERWERKE AG.
 XX
 PI Muller M, Gissmann L;
 XX
 DR WPI; 1991-304643/42.
 XX
 PT Sero-active epitope(s) of human papilloma-virus 16 proteins - for use
 PT vaccines and diagnosis.
 XX
 PS Claim 1; Page 11; 15pp; English.
 XX
 CC This is one of two seroreactive epitopes identified from HPV16 protein
 CC E4. Peptides containing this epitope are also claimed and can be used to
 CC generate antibodies to HPV. See also AAR14288-R14302 and AAQ14168-Q14171
 XX
 SQ Sequence 11 AA;
 Query Match 46.5%; Score 66; DB 2; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.023;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 IPKPSWAPKK 13
 Db 1 IPKPSWAPKK 11
 RESULT 9
 ADE54638
 ID ADE54638 standard; protein; 1053 AA.
 XX
 AC ADE54638;

XX 29-JAN-2004 (first entry)
DT Human Protein BAA20798, SEQ ID NO 443.
DE Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX Homo sapiens.
OS WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX (GHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX GENBANK; BAA20798.
XX New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.
XX Claim 1; Page; 1017pp; English.
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1053 AA;
XX Query Match 43.3%; Score 61.5; DB 7; Length 1053;
XX Best Local Similarity 33.3%; Pred. No. 11;
XX Matches 11; Conservative 7; Mismatches 2; Indels 13; Gaps 1;
XX 2 PIPKPSWAPKXK-----RRLSDQ 21
Db 890 PIPKPSWAPKXK-----RRLSDQ 922

RESULT 10
ADE54636
ID ADE54636 standard; protein; 1553 AA.
XX AC ADE54636;
XX 29-JAN-2004 (first entry)
XX DE Rat Protein AAB66703, SEQ ID NO 441.
XX KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX OS Rattus norvegicus.
XX WO2003016475-A2.
XX 27-FEB-2003.
XX 14-AUG-2002; 2002WO-US025765.
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX (GHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
XX GENBANK; AAB66703.
XX New composition comprising two or more isolated polypeptides, useful for
XX preparing a medicament for treating pain in an animal.
XX Claim 1; Page; 1017pp; English.
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1553 AA;
XX Query Match 43.3%; Score 61.5; DB 7; Length 1553;
XX Best Local Similarity 33.3%; Pred. No. 17;
XX Matches 11; Conservative 7; Mismatches 2; Indels 13; Gaps 1;
XX 2 PIPKPSWAPKXK-----RRLSDQ 21

[illegible]

[illegible]

PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US004927.
 XX
 PR 28-FEB-2000; 2000US-00515126.
 PR 18-MAY-2000; 2000US-00577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI82954.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
 PT and treating e.g. leukemia, inflammation and immune disorders.
 XX
 PS Claim 20; SEQ ID NO 16915; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 105 AA;
 Query Match 39.4%; Score 56; DB 4; Length 105;
 Best Local Similarity 81.8%; Pred. No. 5.8;
 Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 4 PKPSWAPKXH 14
 DB 82 PCPSWAPKTH 92
 Search completed: May 27, 2004, 16:18:27
 Job time : 72.2308 secs